Geometric Continuity: A Parametrization Independent Measure of Continuity for Computer Aided Geometric Design

by

Anthony D. DeRose

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Computer Science

University of California
Berkeley, California

August, 1985

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
AUG 1985		**00-00-1985 to 00-00-1985**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Geometric Continuity: A Parametrization Independent Measure of Continuity for Computer Aided Geometric Design	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California at Berkeley,Department of Electrical Engineering and Computer Sciences,Berkeley,CA,94720**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Parametric spline curves and surfaces are typically constructed so that some number of derivatives match where the curve segments or surface patches abut. If derivatives up to order n are continuous, the segments or patches are said to meet with C^ n, or nth order parametric continuity. It has been shown previously that parametric continuity is sufficient, but not necessary, for geometric smoothness. The geometric measures of unit tangent and curvature vectors for curves (objects of parametric dimension one), and tangent plane and Dupin indicatrix for surfaces (objects of parametric dimension two), have been used to define first and second order geometric continuity. These measures are intrinsic in that they are independent of the parametrizations used to describe the curve or surface. In this work, the notion of geometric continuity as a parametrization independent measure is extended for arbitrary order n (G^n), and for objects of arbitrary parametric dimension p. Two equivalent characterizations of geometric continuity are developed: one based on the notion of reparametrization, and one based on the theory of differentiable manifolds. From the basic definitions, a set of necessary and sufficient constraint equations is developed. The constraints (known as the Beta constraints) result from a direct application of the univariate chain rule for curves and the bivariate chain rule for surfaces. In the spline construction process the Beta constraints provide for the introduction of freely selectable quantities known as shape parameters. For polynomial splines, the use of the Beta constraints allows greater design flexibility through the shape parameters without raising the polynomial degree. The approach taken is important for several reasons. First, it generalizes geometric continuity to arbitrary order for both curves and surfaces. Second, it shows the fundamental connection between geometric continuity of curves and that of surfaces. Third, due to the chain rule derivation, constraints of any order can be determined more easily than using derivations based exclusively on geometric measures. Finally, a firm connection is established between the theory of differentiable manifolds and the use of parametric splines in computer aided geometric design.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**146**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

GEOMETRIC CONTINUITY:
A PARAMETRIZATION INDEPENDENT MEASURE OF CONTINUITY FOR COMPUTER AIDED GEOMETRIC DESIGN

Anthony D. DeRose

FOR MY MOTHER

Acknowledgments

I would like to thank my thesis advisor, Brian Barsky, for providing a sometimes hectic, but always exciting, research environment. Through him, many opportunities have been presented to me that would have otherwise been impossible. I would also like to thank Beresford Parlett for serving as the ever-present "outside member", and for enduring several afternoons of obscure questions.

Special thanks is in order to Ron Goldman, the third member of my committee. From the beginning, Ron has treated this research with astounding care and enthusiasm, often providing deep, insightful comments. It must be rare for a graduate student to find a committee member willing to spend the kind of time and effort that Ron has devoted to this dissertation, and I am indeed grateful for his participation.

The arduous progression from first year status, through prelims, quals, and finally the dissertation, has actually been quite pleasurable, due in large part to an eclectic group of friends, including: Gregg Foster, Mark Hill, Susan Eggers, Steve Upstill, Prabhakar Ragde, John Gross, Ken Fishkin, Lindy Foster, and my roommates for the last two years, Garth Gibson and Pounce.

I'd also like to thank Mom, Dad and my sister Dianne, for believing me when I said I'd someday finish my P-H-D and get a real J-O-B. Finally, there's Cindy Babuska — the gal that has had to put up with me through this ordeal. Lord knows it hasn't been easy for her, but she didn't even complain when we had to cancel our Mexican vacation.

This work was supported in part by the Defense Advanced Research Projects Agency under contract number N00039-82-C-0235, the National Science Foundation under grant number ECS-8204381, the State of California under a Microelectronics Innovation and Computer Research Opportunities grant, and a Shell Doctoral Fellowship.

Contents

Geometric Continuity: A Parametrization Independent Measure of Continuity for Computer Aided Geometric Design

Anthony D. DeRose

ABSTRACT

Parametric spline curves and surfaces are typically constructed so that some number of derivatives match where the curve segments or surface patches abut. If derivatives up to order n are continuous, the segments or patches are said to meet with C^n, or n^{th} order *parametric continuity*. It has been shown previously that parametric continuity is sufficient, but not necessary, for geometric smoothness.

The geometric measures of *unit tangent* and *curvature* vectors for curves (objects of parametric dimension one), and *tangent plane* and *Dupin indicatrix* for surfaces (objects of parametric dimension two), have been used to define first and second order *geometric continuity*. These measures are *intrinsic* in that they are independent of the parametrizations used to describe the curve or surface. In this work, the notion of geometric continuity as a parametrization independent measure is extended for arbitrary order n (G^n), and for objects of arbitrary parametric dimension p. Two equivalent characterizations of geometric continuity are developed: one based on the notion of *reparametrization*, and one based on the theory of *differentiable manifolds*.

From the basic definitions, a set of necessary and sufficient constraint equations is developed. The constraints (known as the *Beta constraints*) result from a direct application of the univariate chain rule for curves and the bivariate chain rule for surfaces. In the spline construction process the Beta constraints provide for the introduction of freely selectable quantities known as *shape parameters*. For polynomial splines, the use of the Beta constraints allows greater design flexibility through the shape parameters without raising the polynomial degree.

The approach taken is important for several reasons. First, it generalizes geometric continuity to arbitrary order for both curves and surfaces. Second, it shows the fundamental connection between geometric continuity of curves and that of

surfaces. Third, due to the chain rule derivation, constraints of any order can be determined more easily than using derivations based exclusively on geometric measures. Finally, a firm connection is established between the theory of differentiable manifolds and the use of parametric splines in computer aided geometric design.

1

Introduction

In the early days of computer aided geometric design (CAGD) and computer graphics, it was common to model objects with linear segments or planar polygonal facets. However, polygonal modeling is not well suited to the modeling of smoothly varying objects such as the outline of a character in a typography system, the surface of a ship hull, or the skin of an airplane. To define objects such as these, higher order curve and surface models known *parametric splines* have become popular. Parametric splines are piecewise functions, so care must be taken to "stitch" the curve segments or surface patches together in a "smooth" fashion. It is the fundamental notion of smoothness that this work addresses.

The usual measure of smoothness, known as *parametric continuity*, requires the piecing together of curves and surfaces so that a given number of *parametric derivatives* match at the boundaries between curve segments or surface patches. The *order of continuity* (the number of derivatives that are required to match) is determined by the particular application. Although this generally results in splines that "look" smooth, it is shown in Chapter 2 that parametric continuity can be overly restrictive since it depends upon details of the parametrizations that are irrelevant for many CAGD applications.

To remedy this situation, we define a measure of continuity, known as *geometric continuity,* that is insensitive to changes in these irrelevant details. In other words, geometric continuity is a *parametrization independent* measure of continuity. The definition of geometric continuity (to be given in Chapter 2) is concise and conceptually simple, but it is a definition based on the existence of certain *equivalent parametrizations.* As such, it is a definition that does not lend itself to practical use.

To achieve a practical definition of geometric continuity, a set of constraint equations, known as *Beta constraints*, is derived from the basic definition. These constraints are necessary and sufficient conditions for geometric continuity of the spline. It is shown that the Beta constraints result from a direct application of the *chain rule* for differentiation: the univariate chain rule for curves, the bivariate chain rule for surfaces, the trivariate chain rule for volumes, and so on for objects of higher *parametric dimension.*

Geometric continuity is not only of theoretic interest, but also has practical uses. Due to its parametrization independent nature, a large number of splines that are not parametrically smooth *are* geometrically smooth. For instance, the class of piecewise geometrically continuous polynomial splines of degree d strictly includes the class of piecewise parametrically continuous splines of degree d (see Figure 1.1).

The generality of geometric continuity is reflected in the Beta constraints through the introduction of variable quantities called *shape parameters.* The shape parameters are degrees of freedom that are not available when using parametric continuity. The shape parameters can be made available to a designer in a CAGD environment. If the spline technique is based on the Beta constraints rather than the parametric continuity constraints, the shape parameters can be used to alter the shape of the curve or surface. Experience has shown that shape parameter modification can be a very effective method of shape control [3,4,5]. Moreover, shape parameter modification can be performed independent of the other controls the designer has over shape. Specific examples of this process are given in Chapters 3, 4, and 5.

Previous definitions of parametrization independent measures of continuity were either based on certain fundamental geometric measures, to be discussed in Section 2.2, or, like ours, were based on the existence of equivalent parametrizations. Each of these approaches have had their problems. For instance, the derivation of the Beta constraints from geometric measures is often cumbersome and error prone. Moreover, it seems difficult to define continuity for order higher than two, and if such definitions were to be stated, it is likely that the algebra required to derive the Beta constraints would be prohibitively complex. On the other hand, previous work on existence definitions have failed to derive general Beta constraints for third order and higher. Finally, all previous work has treated curves and surfaces completely separately. Thus, all practical work on parametrization independent measures has been restricted to first and second order continuity for

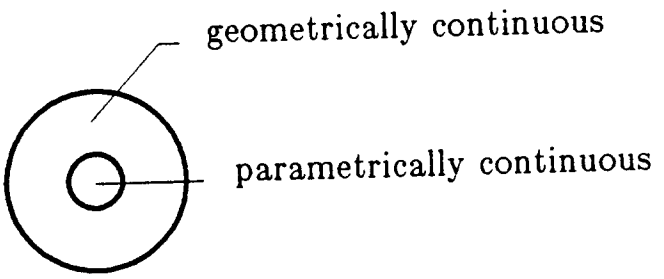

Figure 1.1. *The class of geometrically continuous polynomial splines of a given degree is a strict generalization of the class of parametrically continuous splines of the same degree. The shape parameters are used to explore the larger design space provided by geometric continuity.*

curves and surfaces (see Figure 1.2).

Although first and second order continuity for curves and surfaces is sufficient for many applications, there are situations where higher order continuity or higher parametric dimension is needed. For instance, third order continuous curves and surfaces find application in ship hull design [48], and objects of parametric dimension three and four naturally arise in computer graphics animation [25]. Thus, it is desirable to define and describe geometric continuity for higher order and higher parametric dimension.

This work characterizes geometric continuity for arbitrary order, and for arbitrary parametric dimension. Perhaps more importantly, the extension to higher order continuity and higher parametric dimension is done in a *unified way*, and it is shown that the Beta constraints can be easily derived. The extension is unified in the sense that geometric continuity for curves, surfaces, volumes, etc., are all developed as manifestations of the same underlying theory (see Figure 1.2).

The primary purpose of this work is to characterize the *nature* of geometric continuity by examining the smoothness properties of geometrically continuous splines, providing many alternate definitions of geometric continuity, and proving that the Beta constraints may be easily derived. However, it is not the purpose of this work to completely characterize the *use* of geometric continuity in CAGD. Rather, we present several examples of its use, leaving a comprehensive investigation of the uses of geometric continuity as a topic of future research.

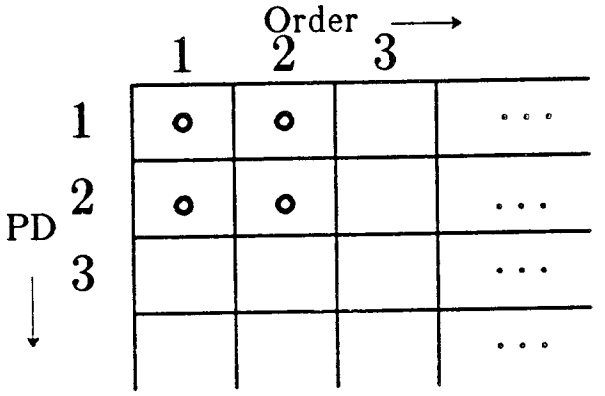

Figure 1.2. *The figure above is a tabular depiction of the "problem space" of geometric continuity. The order of continuity increases to the right, and the parametric dimension increases downward. The first row therefore corresponds to geometric continuity for curves, the second row to surfaces, the third row to volumes, and so on. Circles denote those instances of geometric continuity for which basic definitions have previously been given, and Beta constraints derived. This work fills out all entries of the doubly infinite table in a unified way.*

1.1. Overview

The presentation is logically divided into three parts:

Part I, consisting of Chapter 2, presents an intuitive approach to geometric continuity for curves and surfaces. Virtually all the central results of the theory are developed using plausibility arguments rather than formal proofs. The development is based on the notion of *reparametrization* in conjunction with the chain rule. The relationship between this work and previous work is also discussed.

Part II, consisting of Chapters 3, 4, and 5, discusses some applications of the theory of geometric continuity. In Chapter 3, the use of geometric continuity and the Beta constraints is discussed for parametric curves. Chapter 4 shows that a surface constructed by forming a *tensor product* of geometrically continuous curves, produces a geometrically continuous surface. In Chapter 5, geometrically continuous *triangular* surface techniques are explored. These include the placement of *control vertices* for *Bézier triangles,* and the construction of a new triangular surface technique called the *triangular cubic Beta-spline.* The triangular cubic Beta-spline is a surface technique that possesses one shape parameter and guarantees tangent plane continuity. An evaluation algorithm for triangular cubic

Beta-spline surfaces based on *recursive subdivision* is also developed.

Part III, consisting of Chapter 6, presents a formal development of the concepts and results contained in Part I. This is done by casting the spline construction problem into the language of *differentiable manifolds*. It is shown that splines can be viewed as differentiable deformations of domain manifolds. This view of spline construction is particularly convenient because it allows parametrization independent statements to be made in a natural way. The manifold approach is also important because it firmly establishes the connection between splines in CAGD and differentiable manifold theory, thereby allowing the full power of manifold theory to be brought to bear on problems in CAGD. It is believed that geometric continuity is but one instance of the usefulness of manifold theory in CAGD.

It is suggested that Parts I and II be read by those seeking a high-level understanding of geometric continuity and geometrically continuous spline techniques. Readers interested in a rigorous, in-depth development of the theory of geometric continuity are encouraged to read Part III. Part III may also be of interest to those wishing to use manifold theory in CAGD. However, a warning is in order: Part III assumes a good deal of mathematical sophistication. This method of presentation was chosen so that the rather obscure (albeit powerful) formal development of Part III does not hide the essentially simple ideas and results of geometric continuity presented in Parts I and II.

1.2. Notation and Conventions

- The symbols $\mathbf{Z}$, $\mathbf{Z}_+$, $\Re$, and $\Re_+$ will be used to denote the set of integers, non-negative integers, reals, and non-negative reals, respectively.
- We use a diacritical vector to denote integer tuples such as $\vec{k} = (k_1, k_2, ..., k_n)$. Unless otherwise stated, the components are assumed to be chosen from $\mathbf{Z}_+$. The norm of $\vec{k}$, denoted $|\vec{k}|$, is defined to be the sum of the components of $\vec{k}$.
- Following the notation of Farin [30], it is convenient to denote by $\vec{j}_s$ a tuple whose components are all zero, except for the s^{th} component, which is one.
- The body of definitions, theorems, lemmas, remarks, and examples are set in slanted type to clearly distinguish them from the surrounding text.
- In Parts I and II, we set scalars and scalar-valued functions in Italics; vectors, points, and point-valued functions are set in bold face type. This convention does not apply to Part III however, since no distinction is made between

scalars, and points (or vectors). That is, a scalar is simply a point in $\Re^m$, with $m = 1$.

- Given a function such as $f(x)$, we denote the i^{th} derivative by $f^{(i)}(x)$.
- Given a function such as $g(x, y)$, we use the notation $g^{(i,j)}(x, y)$ to denote the i^{th} partial derivative with respect to x and the j^{th} partial with respect to y. That is,

$$g^{(i,j)}(x, y) = \frac{\partial^{i+j} g}{\partial^i x \partial^j y}.$$

- For brevity, and when no ambiguity can arise, the point of evaluation (x, y) is left off expressions such as $g^{(1,0)}(x, y)$ yielding simply $g^{(1,0)}$.
- Wherever possible, we use the convention that piecewise functions are denoted by upper case letters; the corresponding lower case letter with appropriate subscripts will be used to denote the constituent functions.

2

An Intuitive Approach

This chapter takes an intuitive approach to the development of geometric continuity. The word "intuitive" is meant to suggest that plausibility arguments rather than formal proofs will be used. We begin with some background material, then show that parametric continuity can be overly restrictive for many applications in CAGD. Next we examine previous work dealing with parametrization independent measures of continuity. We then develop geometric continuity of arbitrary order for curves, including a derivation of the univariate Beta constraints. Finally, the notion of geometric continuity is extended to surfaces.

2.1. Introduction

Computer-based modeling requires an unambiguous definition of objects in a form that can be efficiently stored and manipulated. Perhaps the simplest method of definition is *polygonal modeling*. In a polygonal model, objects are described using points in space, called *vertices*. The vertices are then logically connected with *edges*; a closed loop of vertices and edges defines a *face*, and a collection of faces defines an object. Despite their simplicity, polygonal models are not well suited to the modeling of objects composed of curved boundaries and smoothly varying surfaces.

There are many ways to model curves and surfaces in a computer amenable way. Quadric surfaces, implicit equations, solutions of differential equations, and parametric functions are but a few of the possibilities. Due to their relative simplicity and flexibility, we will concentrate on curve and surface definitions based

on parametric functions.

Curves can be defined or *generated* by one variable parametric functions, also known as *univariate parametrizations*. A univariate parametrization is a point-valued function such as $\mathbf{l}(u) = (x(u), y(u))$, where the *domain parameter* u is allowed to range over some interval $[u_0, u_1]$. For a given value of u, the function $\mathbf{l}(u)$ can be thought of as locating a particle in Euclidean two-space. As u is increased over the interval, the particle traverses a path defined by $\mathbf{l}$, tracing out a curve in the process (see Figure 2.1). If $[u_0, u_1]$ is thought of as an oriented line segment, then $\mathbf{l}$ can be thought of as a *deformation* producing an *oriented curve*. One advantage of the parametric representation is that a curve in Euclidean space of arbitrary dimension d can be described by a parametrization $\mathbf{l}(u) = (x_1(u), x_2(u), ..., x_d(u))$.

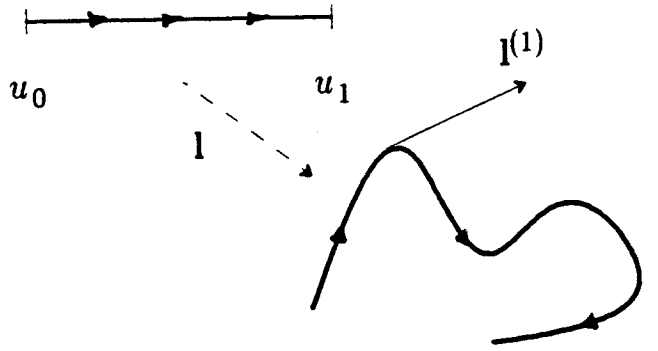

Figure 2.1. *The univariate parametrization* $\mathbf{l}$ *generates an oriented curve by deformation of the oriented line segment* $[u_0, u_1]$.

A *surface patch* in three-space can be defined by a bivariate function such as

$$\mathbf{G}(u, v) = (X(u, v), Y(u, v), Z(u, v)),$$

where u and v are allowed to range over some region D_G of the uv plane (see Figure 2.2). Surface patches in higher dimensions can be described by adding additional component functions. Loosely speaking, a *surface* is a collection of surface patches.

For a curve generated by $\mathbf{l}(u)$, the first derivative vector $\mathbf{l}^{(1)}(u)$ represents the *velocity* of the particle. The velocity is a vector quantity and, as such, contains information about orientation and *rate*, or speed. The second derivative vector

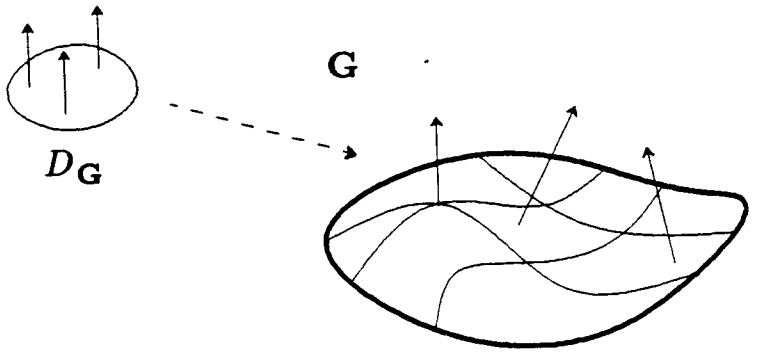

Figure 2.2. *The bivariate parametrization* **G** *deforms the oriented domain* D_G *to generate an oriented surface patch.*

$\mathbf{l}^{(2)}$ represents the *acceleration* of the particle, so it too contains information about the rate (specifically, the change of rate). Thus, a parametrization contains information about the *geometry* (the shape or image of the curve), the orientation, and the rate. Figure 2.3 shows the curves generated by three different parametrizations. The shape of the curves is identical; they differ only in orientation and rate. Curves (a) and (b) have the same orientation at each point, but the rates differ. The curve labeled (c) differs from (a) and (b) in orientation *and* rate. If a curve is defined to be simply the geometry of a parametrization, one would conclude that figures (a), (b), and (c) represent *equivalent* curves. We will refer to this as the *G model* of a curve. Another possibility is to consider the geometry and orientation, which we will call the *GO model.* Using the GO model, one would say that (a) and (b) are equivalent, but (c) is different. The last possibility we will consider is the *GOR model*, where geometry, orientation, and rate are all relevant to the definition of a curve. Using this model, no pair of the curves in Figure 2.3 is equivalent.

Parametric spline curves are typically constructed by stitching together univariate parametric functions, requiring that some number of derivatives match at each *joint* (the points where the curve segments meet). If n derivatives agree at a given joint, the parametrizations there are said to meet with n^{th} *order parametric continuity* (C^n continuity for short).

We maintain that the choice of a particular model for a curve, and hence the choice of how the curve segments are stitched together, should be application dependent. For instance, if a spline is being used to define the motion of an object in an animation system, the GOR model is most appropriate since orientation and

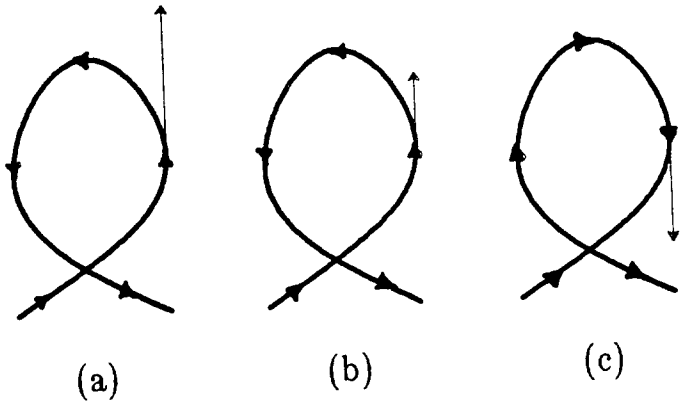

Figure 2.3. *Each of the curves above has the same image; they only differ in orientation and rate. Orientation is indicated by arrowheads and rate is indicated by vectors tangent to the curves.*

rate are important. In this type of application, parametric continuity is required to maintain the smoothness of the rate. In other words, parametric continuity will ensure that the object will *move* smoothly.

However, in CAGD the rate of a parametrization is often unimportant. Consider for example the use of splines to describe numerically-controlled cutters. It may be necessary to specify uniquely the direction of the cutter at each point on the path, but the speed of the cutter may depend upon the hardness of the material being cut. For this type of application, the GO model is most suitable, but parametric continuity is overly restrictive since it places emphasis on irrelevant rate information. The structure provided by orientation can also be useful for surfaces. An oriented surface has a consistently defined normal vector that allows the notions of "top" and "bottom", or "inside" and "outside" to be uniquely defined. This information can often be useful in practice. For instance, a renderer in a computer graphics environment can use orientation information to shade the top of the surface differently from the bottom.

Many other applications in CAGD require only the G model. For instance, if a spline curve is being used to describe the outline of a character in a typography system, only the shape of the outline is relevant. However, it is difficult to develop a useful formalism based only on the geometry of a curve or surface. The difficulty arises because of the *global nature* of the G model. That is, the geometry of a curve or surface cannot be completely characterized by examining local neighborhoods —

geometry is a global property. In fields such as topology and differential geometry it is well known that global statements are difficult and few. Orientation, however, is a local property, so the GO model is also local. The structure provided by orientation allows a local theory of continuity to be developed. Hence, we adopt the GO model and develop an appropriate measure of continuity — one based only on geometry and orientation. We refer to this measure as *geometric continuity*, a term first introduced by Barsky & Beatty [6]. Although it is common to use parametric polynomial parametrizations, the development of geometric continuity we present is valid for an extremely large class of parametrizations to be identified subsequently.

2.2. Previous Work

Many authors have independently defined parametrization independent measures of continuity for first and second order (which we denote by G^1 and G^2, respectively) for curves and/or surfaces using geometric means. For curves, Fowler & Wilson [34], Sabin [52], Manning [47], Faux & Pratt [32], and Barsky [3] each defined first order continuity by requiring that the *unit tangent vectors* agree at the joints. For a curve generated by $\mathbf{l}(u)$, the unit tangent vector at the point $\mathbf{l}(u)$, denoted $\widehat{\mathbf{t}}(u)$, points in the same direction as $\mathbf{l}^{(1)}(u)$, but is required to be of unit length (see Figure 2.4). Thus, $\widehat{\mathbf{t}}(u)$ is defined by

$$\widehat{\mathbf{t}}(u) = \frac{\mathbf{l}^{(1)}(u)}{|\mathbf{l}^{(1)}(u)|}. \tag{2.1}$$

The unit tangent is a parametrization independent characterization of the curve to first order. The requirement of matching unit tangent vectors is therefore a first order parametrization independent measure of continuity for curves (see Figure 2.4).

A second order parametrization independent characterization for curves is provided by the *osculating circle*, or equivalently, the *curvature vector*. Intuitively, the osculating circle and the curvature vector measure the rate at which the curve deviates from its tangent direction. The curvature vector, denoted $\mathbf{k}(u)$, can analytically expressed in terms of derivatives of $\mathbf{l}(u)$ as [3,4]

$$\mathbf{k}(u) = \frac{\mathbf{l}^{(1)}(u) \times \mathbf{l}^{(2)}(u) \times \mathbf{l}^{(1)}(u)}{|\mathbf{l}^{(1)}(u)|^4}. \tag{2.2}$$

The osculating circle is tangent to the curve at $\mathbf{l}(u)$, having a radius that is the reciprocal of the magnitude of the curvature vector, as shown in Figure 2.4. The plane in which the osculating circle lies is called the *osculating plane*.

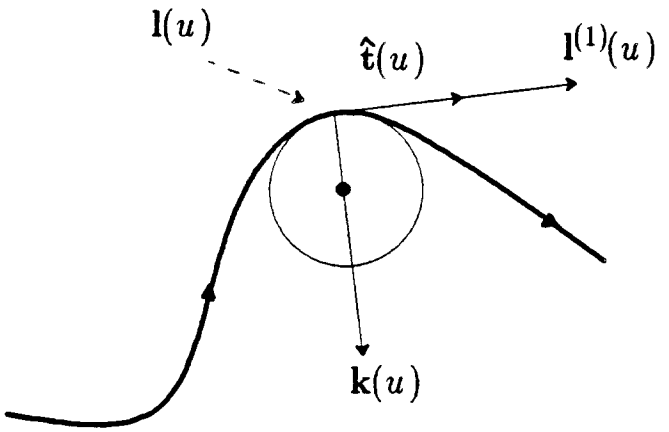

Figure 2.4. *The unit tangent vector, the curvature vector, and the osculating circle at a point* $\mathbf{l}(u)$.

Fowler & Wilson, Sabin, Manning, Faux & Pratt, and Barsky defined second order geometric continuity as continuity of the unit tangent and curvature vectors. Nielson's ν-spline [49] possesses a similar kind of continuity having continuous first derivative and curvature vectors. The geometric measures of unit tangent and curvature essentially ignore the rate information by "normalizing" the parametrization before determining smoothness.

For surfaces, the *tangent plane* is a first order characterization that is parametrization independent, so it is common to require matching of tangent planes for first order geometric continuity (cf. Sabin [53] and Veron *et al* [60]). It is well known in differential geometry that tangent plane continuity is necessary for first order smoothness of the composite surface, but it is not sufficient. We will return to this topic in Section 2.5.1.

The situation for second order continuity between surfaces is even more involved. A second order parametrization independent characterization of surfaces is provided by the *osculating paraboloid.* Equivalent characterizations are provided by the *Dupin indicatrix* and the *second fundamental form.* Intuitively, these objects measure the tendency of a surface to deviate from its tangent plane, in much the same way as the curvature vector measures the deviation of a curve from its tangent line. The osculating paraboloid is perhaps the easiest to describe, so we discuss it first; the Dupin indicatrix and the second fundamental form will be defined in terms of the osculating paraboloid.

The osculating paraboloid at a point $\mathbf{G}(u,v)$ is the quadric surface that best approximates the surface generated by $\mathbf{G}$ in the region of $\mathbf{G}(u,v)$. To describe the osculating paraboloid analytically, it is convenient to set up a coordinate system with its origin at $\mathbf{G}(u,v)$. The x axis is chosen to be in the direction of $\mathbf{G}^{(1,0)}(u,v)$, the y axis is chosen to be in the direction of $\mathbf{G}^{(0,1)}(u,v)$, and the z axis is chosen to be in the direction of the unit normal vector given by

$$\hat{\mathbf{N}} = \frac{\mathbf{G}^{(1,0)} \times \mathbf{G}^{(0,1)}}{|\mathbf{G}^{(1,0)} \times \mathbf{G}^{(0,1)}|}, \tag{2.3}$$

as shown in Figure 2.5.

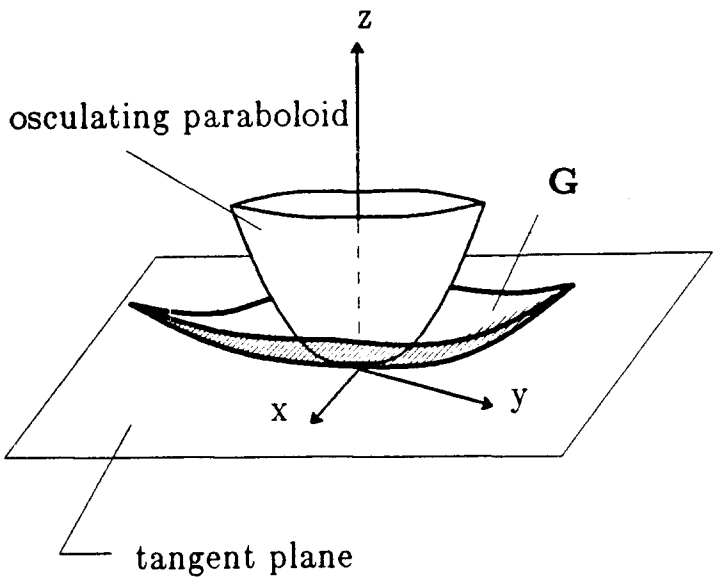

Figure 2.5. *The osculating paraboloid for a surface generated by a parametrization* $\mathbf{G}(u,v)$ *is shown above. The local coordinate system has its origin at* $\mathbf{G}(u,v)$*, the* x *axis along* $\mathbf{G}^{(1,0)}(u,v)$*, the* y *axis along* $\mathbf{G}^{(0,1)}(u,v)$*, and the* z *axis along the normal direction.*

In this coordinate system, the osculating paraboloid is described by the quadratic equation (cf. DoCarmo [26])

$$z = \frac{1}{2}(Lx^2 + 2Mxy + Ny^2) \tag{2.4}$$

where the coefficients L, M, and N are given by

$$\begin{aligned} L &= \mathbf{G}^{(2,0)} \cdot \hat{\mathbf{N}} \\ M &= \mathbf{G}^{(1,1)} \cdot \hat{\mathbf{N}} \\ N &= \mathbf{G}^{(0,2)} \cdot \hat{\mathbf{N}} \end{aligned} \tag{2.5}$$

where $\cdot$ denotes vector dot product.

The coefficients L, M, and N are the components of the 2-tensor called the second fundamental form (cf. Faux & Pratt [32], or Synge & Schild [59]), and the curve formed by intersecting the osculating paraboloid with the either the $z = +1/2$ or $z = -1/2$ plane is the Dupin indicatrix (cf. DoCarmo [26]).

Veron *et al* [60] and Kahmann [44] require continuity of tangent plane and Dupin indicatrix for second order geometric continuity.

2.3. Reparametrization and the Chain Rule

Although the geometric approaches described in Section 2.2 are convenient and intuitive for first and second order continuity, a more algebraic development is better suited for the extension to continuity of higher order. The approach we take is based on *reparametrization* — the process of obtaining a new parametrization given an old one. In the GO model, reparametrization may change rate, but not geometry or orientation. By allowing reparametrization before making a determination of continuity, the rate aspects of parametrizations may be ignored. Alternately stated, our approach is based on the following simple principle:

P1: Don't base continuity on the parametrizations at hand; reparametrize, if necessary, to obtain parametrizations that meet with parametric continuity. If this can be done, the original parametrizations must also meet smoothly, at least in a geometric sense.

The above concept is not a new one; similar principles have been discussed by Farin [27] and Veron *et al* [60]. What is new is the use of the principle to construct constraint equations (known as *Beta constraints*) that are necessary and sufficient for geometric continuity of arbitrary order for both curves and surfaces. *

The Beta constraints generalize the parametric continuity constraints through the introduction of freely variable quantities called *shape parameters*. Once the

* Goodman [37] and Ramshaw [50] have independently derived the univariate Beta constraints from the univariate chain rule.

Beta constraints are determined for a given order of continuity, they may be used in place of the parametric continuity constraints when building splines, thereby yielding increased flexibility. For instance, if the C^2 constraints are replaced with the G^2 constraints in the uniform cubic B-spline [51], the cubic Beta-spline results [3,4]. The cubic Beta-spline, discussed in Section 3.3, is an *approximating* spline technique that possesses two shape parameters; a class of splines containing *interpolating* members is described in DeRose & Barsky [23] (see Section 3.4). Faux & Pratt [32], Farin [27], Fournier & Barsky [33], and Ramshaw [50] use the extra freedom allowed by geometric continuity to place *Bézier control vertices* (see Section 3.2).

An important aspect of the geometrically continuous polynomial techniques mentioned above is that the additional flexibility of geometric continuity can be added without increasing the degree of the polynomials. This is particularly important for algorithms that manipulate the spline. For example, the complexity of Sederberg's algorithm [55] for intersecting two polynomial curves of degree d grows at least as fast as d^3. Substantial savings can be realized by minimizing the degree of the polynomials involved.

In the remainder of this chapter, we extend the notion of geometric continuity to arbitrary order n (G^n) and show (in a nonrigorous way) that the derivation of the Beta constraints results from a straightforward use of the univariate chain rule for curves and the bivariate (two variable) chain rule for surfaces. For a more complete treatment, see Chapter 6 where geometric continuity is characterized in another, but completely equivalent, way using the theory of differentiable manifolds.

2.4. Geometric Continuity for Curves

A univariate parametrization is said to be *regular* if the first derivative vector does not vanish. It is well known from differential geometry (cf. DoCarmo [26]) that regularity is, in general, essential for the smoothness of the resulting curve (see Figure 2.6). We therefore restrict the discussion to parametrizations that are regular. We also make the restriction that parametrizations are infinitely differentiable (C^∞), but no restriction on the dimension of the parametrization is made. Thus, all the results in this section hold for curves in Euclidean space of arbitrary dimension.

We begin the study of geometric continuity for curves by examining the reparametrization process. Two parametrizations are said to be *GO-equivalent*

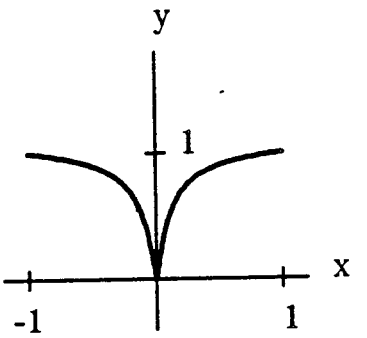

Figure 2.6. *Consider the parametrization* $\mathbf{l}(u) = (u^3, u^2), u \in [-1, 1]$ *shown above. Even though* $\mathbf{l}(u)$ *is infinitely differentiable everywhere on* $[-1, 1]$*, the curve does not have a continuous unit tangent at* $(0, 0)$*, the point on the curve corresponding to* $u = 0$*. This behavior is possible because the derivative vanishes at* $u = 0$*, and thus the parametrization is irregular at* $u = 0$*.*

if they have the same geometry and orientation in the neighborhood of each point. As a consequence of the *Inverse Function Theorem* (see Theorem 6.3, Section 6.3), given a parametrization $\mathbf{l}$, all GO-equivalent parametrizations may be obtained by *functional composition*. More specifically, if $\mathbf{l}(u)$ and $\tilde{\mathbf{l}}(\tilde{u})$ are GO-equivalent, then they are related by $\tilde{\mathbf{l}}(\tilde{u}) = \mathbf{l}(u(\tilde{u}))$, for some appropriately chosen differentiable *change of parameter* $u(\tilde{u})$ (see Figure 2.7).

Since $\mathbf{l}$ and $\tilde{\mathbf{l}}$ must have the same orientation, u must be an increasing function of $\tilde{u}$, implying that u must satisfy the *orientation preserving condition* $u^{(1)} > 0$. Intuitively, $u(\tilde{u})$ deforms the interval $[\tilde{u}_0, \tilde{u}_1]$ into the interval $[u_0, u_1]$ without reversing the orientation of the segment $[\tilde{u}_0, \tilde{u}_1]$. This in turn implies that $\mathbf{l}$ and $\tilde{\mathbf{l}}$ will have the same geometry and orientation, but they may differ in rate. We now give a more precise definition of G^n continuity:

Definition 2.1: *Let* $\mathbf{l}(u), u \in [u_0, u_1]$ *and* $\mathbf{r}(t), t \in [t_0, t_1]$ *be two regular* C^∞ *parametrizations such that* $\mathbf{l}(u_1) = \mathbf{r}(t_0) = \mathbf{J}$ *(see Figure 2.8). These parametrizations meet with* G^n *continuity at* $\mathbf{J}$ *if there exist GO-equivalent parametrizations* $\tilde{\mathbf{l}}(\tilde{u})$ *and* $\tilde{\mathbf{r}}(\tilde{t})$ *that meet with* C^n *continuity at* $\mathbf{J}$*.*

Definition 2.1 is simply a restatement of principle P1. In practice one cannot examine all GO-equivalent parametrizations in an effort to find two that meet with parametric continuity. However, it is possible to find conditions on $\mathbf{l}$ and $\mathbf{r}$ that are necessary and sufficient for the *existence* of GO-equivalent parametrizations that meet with parametric continuity.

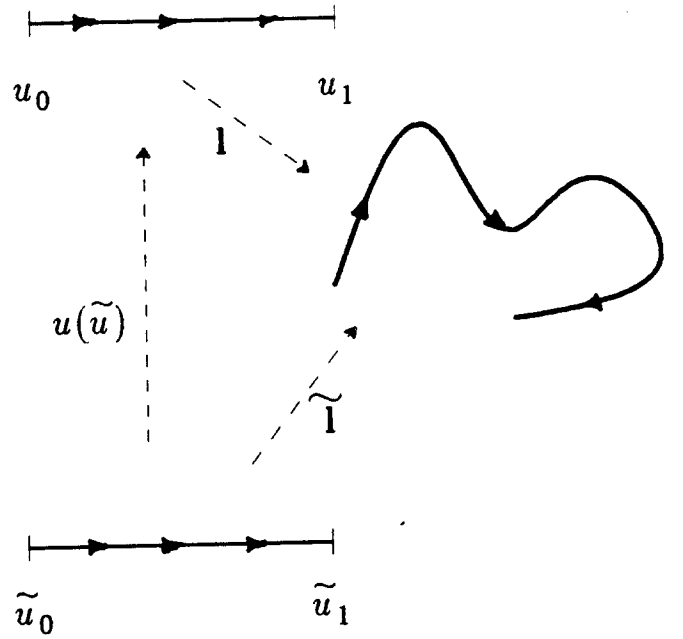

Figure 2.7. *The GO-equivalent parametrizations* $\mathbf{l}$ *and* $\tilde{\mathbf{l}}$ *are related by the change of parameter* $u(\tilde{u})$.

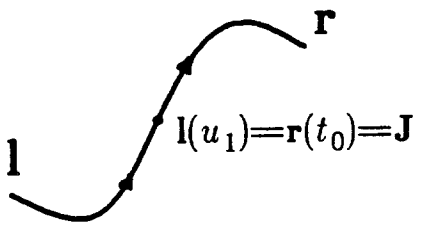

Figure 2.8. *The parametrizations* $\mathbf{l}(u)$ *and* $\mathbf{r}(t)$ *meet at the common point* $\mathbf{J}$.

Although Definition 2.1 suggests that both $\mathbf{l}$ and $\mathbf{r}$ need to be reparametrized, it is possible to show that Definition 2.1 holds if and only if there exists a $\tilde{\mathbf{l}}$ that meets $\mathbf{r}$ with parametric continuity. In other words, only one of the parametrizations needs to be reparametrized to determine smoothness.

We will ultimately be interested in the derivative properties of $\tilde{\mathbf{l}}$. The *univariate chain rule* allows us to express derivatives of $\tilde{\mathbf{l}}$ in terms of the derivatives

of $\mathbf{l}$ and u. For example, the first derivative is given by

$$\begin{aligned}\tilde{\mathbf{l}}^{(1)} = \frac{d\tilde{\mathbf{l}}}{d\tilde{u}} &= \frac{d\mathbf{l}(u(\tilde{u}))}{d\tilde{u}} \\ &= \frac{du}{d\tilde{u}}\frac{d\mathbf{l}}{du} \\ &= u^{(1)}\mathbf{l}^{(1)}.\end{aligned} \tag{2.6}$$

In general, the i^{th} derivative of $\tilde{\mathbf{l}}$ can be written as a function, call it $\mathcal{CR}_i$ (short for Chain Rule), of the first i derivatives of u and $\mathbf{l}$. That is,

$$\tilde{\mathbf{l}}^{(i)} = \mathcal{CR}_i(\mathbf{l}^{(1)}, \cdots, \mathbf{l}^{(i)}, u^{(1)}, \cdots, u^{(i)}). \tag{2.7}$$

We are actually interested in $\tilde{\mathbf{l}}^{(i)}$ evaluated at its *right parametric endpoint* $\tilde{u}_1$. Thus, derivatives of $\mathbf{l}$ and u must also be evaluated at their right parametric endpoints:

$$\begin{aligned}\tilde{\mathbf{l}}^{(i)}(\tilde{u}_1) = \mathcal{CR}_i(&\mathbf{l}^{(1)}(u_1), \cdots, \mathbf{l}^{(i)}(u_1), \\ &u^{(1)}(\tilde{u}_1), \cdots, u^{(i)}(\tilde{u}_1)).\end{aligned} \tag{2.8}$$

Since u is a scalar function, evaluating one of its derivatives results in a real number. In particular, let $\beta_j = u^{(j)}(\tilde{u}_1)$, $j = 1, ..., i$. Equation (2.8) then becomes

$$\tilde{\mathbf{l}}^{(i)}(\tilde{u}_1) = \mathcal{CR}_i(\mathbf{l}^{(1)}(u_1), \cdots, \mathbf{l}^{(i)}(u_1), \beta_1, \cdots, \beta_i). \tag{2.9}$$

The orientation preserving quality of u implies that $\beta_1 > 0$.

We are now in a position to state the primary result of geometric continuity for curves. Recall that $\mathbf{l}$ and $\mathbf{r}$ meet with G^n continuity if $\mathbf{l}$ can be reparametrized to $\tilde{\mathbf{l}}$ so that derivatives of $\mathbf{r}$ and $\tilde{\mathbf{l}}$ agree. That is, we require that

$$\mathbf{r}^{(i)}(t_0) = \tilde{\mathbf{l}}^{(i)}(\tilde{u}_1), \qquad i = 1, ..., n. \tag{2.10}$$

Positional continuity is implicitly assumed (see Figure 2.8). Substituting equation (2.9) into (2.10) yields

$$\mathbf{r}^{(i)}(t_0) = \mathcal{CR}_i(\mathbf{l}^{(1)}(u_1), \cdots, \mathbf{l}^{(i)}(u_1), \beta_1, \cdots, \beta_i), \qquad i = 1, ..., n. \tag{2.11}$$

The constraints resulting from equation (2.11) are the *univariate Beta constraints* and the numbers $\beta_1, ..., \beta_n$ are the *shape parameters*. The above discussion is not a proof that the Beta constraints are necessary and sufficient conditions for geometric continuity, but such a proof can be constructed (see Section 6.9, or, for a more elementary proof, see Barsky & DeRose [6]). For completeness, we formally state the theorem for curves:

Theorem 2.1: *Let* $\mathbf{l}(u)$ *and* $\mathbf{r}(t)$ *be as in Definition 2.1. They meet with* G^n *continuity at* $\mathbf{J}$ *if and only if there exist real numbers* $\beta_1, ..., \beta_n$, *with* $\beta_1 > 0$, *such that equations (2.11) are satisfied.*

Theorem 2.1 states that if equations (2.11) are satisfied for any choice of the β's, subject to $\beta_1 > 0$, then the coincident curve segments will meet with G^n continuity. For instance, the Beta constraints for G^4 continuity between $\mathbf{l}$ and $\mathbf{r}$ are

$$\begin{aligned}
\mathbf{r}^{(1)}(t_0) &= \beta_1 \mathbf{l}^{(1)}(u_1) \\
\mathbf{r}^{(2)}(t_0) &= \beta_1^2 \mathbf{l}^{(2)}(u_1) + \beta_2 \mathbf{l}^{(1)}(u_1) \\
\mathbf{r}^{(3)}(t_0) &= \beta_1^3 \mathbf{l}^{(3)}(u_1) + 3\beta_1\beta_2 \mathbf{l}^{(2)}(u_1) + \beta_3 \mathbf{l}^{(1)}(u_1) \\
\mathbf{r}^{(4)}(t_0) &= \beta_1^4 \mathbf{l}^{(4)}(u_1) + 6\beta_1^2\beta_2 \mathbf{l}^{(3)}(u_1) \\
&\quad + (4\beta_1\beta_3 + 3\beta_2^2)\mathbf{l}^{(2)}(u_1) + \beta_4 \mathbf{l}^{(1)}(u_1).
\end{aligned} \tag{2.12}$$

The discussion leading from equation (2.8) to equation (2.9) suggests that the i^{th} Beta constraint is determined by repeated application of the chain rule, followed by evaluation at the right parametric endpoint. There is, however, an easier way to derive the Beta constraints. The method is based on the observation that $\mathcal{CR}_i$ is obtained by differentiation of $\mathcal{CR}_{i-1}$, suggesting that the i^{th} Beta constraint can be obtained by "differentiating" the $i-1^{\text{st}}$ order constraint. Differentiating in the normal way makes little sense because the Beta constraints are not functions. However, recall that β_i results from the evaluation of $u^{(i)}$, and that $u^{(i+1)}$ results from differentiating of $u^{(i)}$; so in some sense, β_{i+1} results from differentiation of β_i. More specifically, consider the derivative of $(u^{(i)})^k$:

$$\frac{d(u^{(i)})^k}{d\tilde{u}} = k\,(u^{(i)})^{k-1}\,u^{(i+1)} \tag{2.13}$$

which can be interpreted at the right parametric endpoint as

$$\frac{d\beta_i^k}{d\tilde{u}} = k\beta_i^{k-1}\beta_{i+1}. \tag{2.14}$$

Similarly, using the chain rule, the derivative of $\mathbf{l}^{(i)}$ with respect to $\tilde{u}$ is

$$\frac{d\mathbf{l}^{(i)}}{d\tilde{u}} = u^{(1)}\mathbf{l}^{(i+1)} \tag{2.15}$$

which can be interpreted at the right parametric endpoint as

$$\frac{d\mathbf{l}^{(i)}(u_1)}{d\tilde{u}} = \beta_1 \mathbf{l}^{(i+1)}(u_1). \tag{2.16}$$

Using the heuristic rules (2.14) and (2.16) together with the heuristic

$$\frac{d\mathbf{r}^{(i)}(t_0)}{d\tilde{u}} = \mathbf{r}^{(i+1)}(t_0) \tag{2.17}$$

and the product rule for differentiation, the i^{th} order constraint can be obtained by differentiating the $i-1^{\text{st}}$ order constraint with respect to $\tilde{u}$. One can easily verify that equations (2.12) result from these rules.

It is important to check that we have indeed generalized geometric continuity by showing that our definitions reduce to the previous geometry-based definitions of unit tangent and curvature vector continuity. This is easily done by noting that the first two equations of (2.12) are identical to the constraints resulting from a geometric derivation using unit tangent and curvature vector continuity [4,47]. Thus, our approach reduces to previous definitions of G^1 and G^2 continuity for curves. In Section 6.10, it is shown that the Beta constraints for n^{th} order continuity are equivalent to requiring continuity of the first n derivatives with respect to *arc length*.

One way the Beta constraints can be used in practice is to allow the designer to input the values of the shape parameters, perhaps using some graphical device such as a mouse, or analog devices such as continuous-turn dials. The system software then uses internal spline mathematics (to be described in Chapters 3, 4, and 5) to alter the parametrizations of the curve segments so that the Beta constraints are satisfied. Changing the shape parameters will change the shape of the curve or surface, independent of the other controls the designer has over the shape, but the shape change always occurs so as to maintain geometric continuity (see Figure 2.9). As mentioned earlier, the Beta constraints can also be used to govern the positioning of Bézier vertices. This process is described in Section 3.2.

Before moving on to surfaces, it is important to point out that G^2 continuity admits three dimensional curves that change osculating planes suddenly, as shown in Figure 2.10. This behavior can only occur if the curve is not planar *and* the curvature vector vanishes at the joint. If this type of behavior is considered undesirable, more stringent constraints are required for curves in 3-space when the curvature vector vanishes.

2.5. Continuity of Surfaces

In this section, we extend the notion of geometric continuity to surfaces. Since care was taken in Section 2.4 not to base the development of geometric continuity

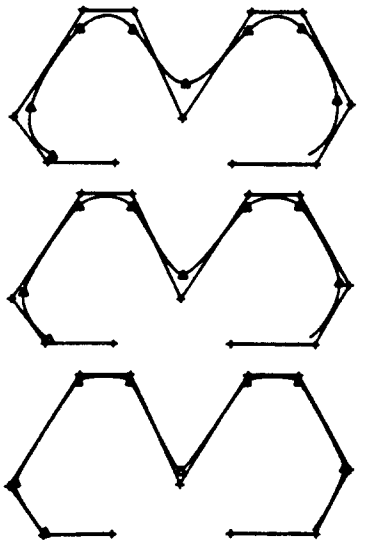

Figure 2.9. *This sequence of curves shows the effect of changing a shape parameter in the cubic Beta-spline technique, to be described in more detail in Section 3.3. Triangles denote the joints between the curve segments. Each of the curves satisfies the G^2 constraints, and all have $\beta_1 = 1$; only β_2 differs from curve to curve. The top curve has $\beta_2 = 0$, the middle curve has $\beta_2 = 5$, and the bottom curve has $\beta_2 = 20$.*

on concepts (such as arc length) that don't apply to surfaces, the machinery developed for univariate parametrizations can be extended readily to bivariate parametrizations.

In Section 2.4, a restriction to regular parametrizations was made. The same restriction is made here so that a unique orientation can be assigned to each point of the surface patch. A bivariate parametrization $\mathbf{G}(u, v)$ is said to be *regular* if the first order partials $\mathbf{G}^{(1,0)}(u, v)$ and $\mathbf{G}^{(0,1)}(u, v)$ are linearly independent for all (u, v). Among other things, regularity guarantees that a well defined unit normal vector can be computed from equation (2.3) at each point.

In Section 2.1, it was seen that univariate parametrizations contain information about geometry, orientation, and rate. The same is true of bivariate parametrizations. Orientation can be defined by treating the domain D_G as an *oriented plane* having a "top side" and a "bottom side." $\mathbf{G}$ can then be thought of as deforming the oriented plane to produce an oriented, or two-sided, surface patch (see Figure 2.2). Rate information enters through the magnitudes (and changes thereof) of the partial derivatives of the parametrization. We can therefore speak

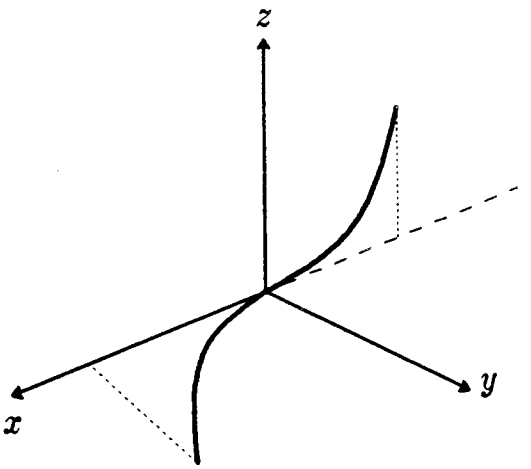

Figure 2.10. *The curve drawn above is defined by* $(u, 0, u^3)$*, for* $u < 0$*, and* $(u, u^3, 0)$ *for* $u \geq 0$*. One can verify that these parametrizations are regular, and have continuous unit tangent and curvature vectors at* $u = 0$*; hence, the parametrizations meet with* G^2 *continuity. However, because the curvature vectors vanish at* $u = 0$*, the osculating plane is allowed to change suddenly. In the example above, when* $u < 0$ *the curve is entirely contained in the* (x, z) *plane, and when* $u \geq 0$ *the curve is entirely contained in the* (x, y) *plane.*

of the G, GO, and GOR models of surfaces. Just as for curves, the use of a particular model should be application dependent. We will adopt the GO model for two reasons: first, orientation is necessary in applications such as rendering where the two-sidedness of surfaces is important, and second, it is difficult to develop a useful formalism without the local structure provided by orientation.

As might be expected, stitching surface patches together is somewhat more involved than curves, so before delving into a detailed discussion of continuity, let us back up and reexamine the description of a surface patch by a bivariate parametrization. In particular, we wish to define the term "parameter space", and make clear the distinct nature of the parameter spaces of the patches that are to be sewn together. Similar care could have been taken with curves described by univariate parametrizations, and strictly speaking, this should have been done in Section 2.4. However, continuity between curve segments can be discussed without carefully maintaining the distinction between parameter spaces, so for simplicity we chose to treat curves with a certain amount of abandon.

Let $\mathbf{G}(u, v)$ be a bivariate parametrization from a domain D_G into Euclidean

three-space. More precisely, D_G is a subset of a Euclidean two-space E_G, with u and v referring to a (not necessarily Cartesian) coordinate system on E_G, and the range of $\mathbf{G}$ is contained in a Euclidean three-space E (see Figure 2.11). Using standard functional notation, we write $\mathbf{G} : D_G \subset E_G \to E$, or, when the space containing D_G is understood, we write $\mathbf{G} : D_G \to E$. E_G is called the *parameter space* of $\mathbf{G}$, and E is called the *image space* of $\mathbf{G}$.

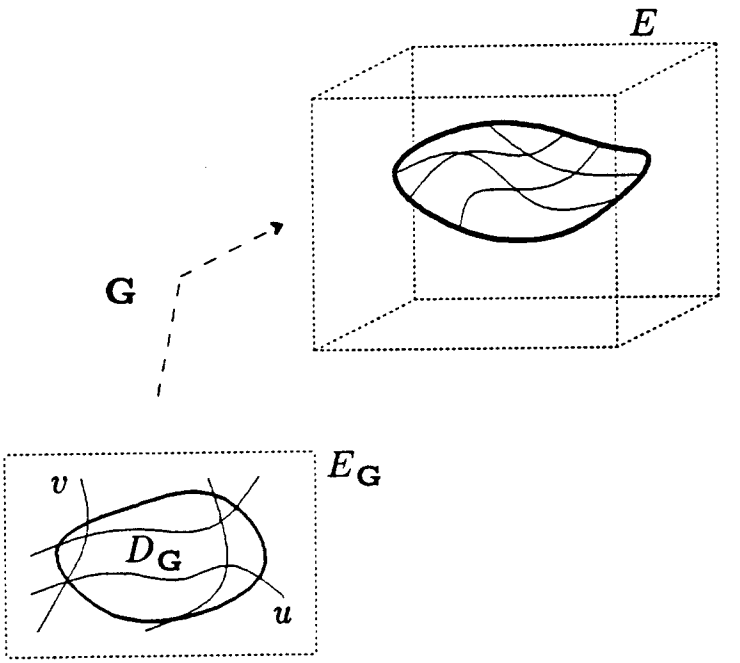

Figure 2.11. *The domain of* $\mathbf{G}$ *is a subset of a Euclidean two-space* E_G*, with the image of* $\mathbf{G}$ *being contained in a Euclidean three-space* E.

In the remainder of this chapter we consider the problem of stitching two patches together. A description of the most general case where a collection of surface patches are stitched together to form a smooth spline surface requires a considerable amount of mathematical machinery, so its exposition is relegated to Chapter 6.

2.5.1. Parametric Continuity for Surface Patches

To begin a study of continuity between two surface patches, consider the situation depicted in Figure 2.12 where $\mathbf{F} : D_F \subset E_F \to E$, having parameters (s,t), and $\mathbf{G} : D_G \subset E_G \to E$, having parameters (u,v), are two regular C^∞

parametrizations meeting with positional continuity along a boundary curve γ. E_F and E_G are initially assumed to be entirely separate, unrelated parameter spaces. If $\mathbf{p}$ is a point on γ, the *preimage* of $\mathbf{p}$ in D_F is the point $(s_\mathbf{p}, t_\mathbf{p})$ that has $\mathbf{p}$ as an image point, that is, $\mathbf{p} = \mathbf{F}(s_\mathbf{p}, t_\mathbf{p})$. Although it is possible for more than one point in D_F to be mapped to $\mathbf{p}$, we currently assume that $\mathbf{F}$ is one-to-one and relax this restriction in Chapter 6. Similarly, let $(u_\mathbf{p}, v_\mathbf{p})$ be the preimage of $\mathbf{p}$ in D_G ($\mathbf{G}$ is also assumed to be 1-1). If the point $\mathbf{p}$ is thought of as being a function of a parameter w, then the boundary curve γ is generated by a univariate parametrization $\mathbf{p}(w)$. For convenience, we assume that the parametrization $\mathbf{p}(w)$ is chosen so that γ is generated when w varies on $[0,1]$; that is, $\mathbf{p} : [0,1] \rightarrow E$. We also assume that γ is smooth in the sense that $\mathbf{p}(w)$ is a regular C^∞ parametrization.

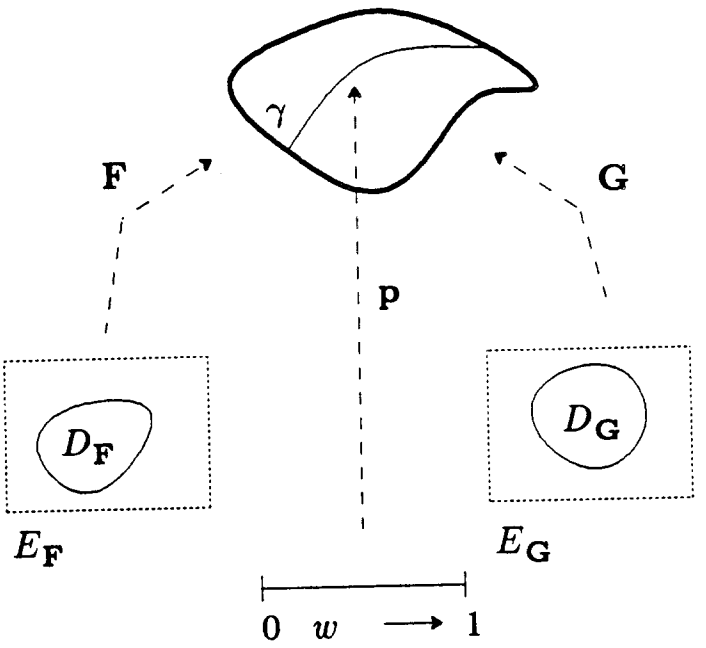

Figure 2.12. *The surface patches generated by the parametrizations* $\mathbf{F}$ *and* $\mathbf{G}$ *meet with positional continuity along a boundary curve* γ*, with* γ *being generated by a parametrization* $\mathbf{p}(w)$*,* $w \in [0,1]$.

The boundary curve γ has a preimage curve γ_F in D_F and a preimage curve γ_G in D_G, as shown in Figure 2.13. The preimage curve γ_F is generated by a parametrization $\mathbf{p}_F(w) = (s_\gamma(w), t_\gamma(w))$, so $\mathbf{p}(w) = \mathbf{F}(s_\gamma(w), t_\gamma(w))$. Similarly,

the preimage curve γ_G is generated by $\mathbf{p}_G(w) = (u_\gamma(w), v_\gamma(w))$, implying that $\mathbf{p}(w) = \mathbf{G}(u_\gamma(w), v_\gamma(w))$, which in turn implies that

$$\mathbf{F}(s_\gamma(w), t_\gamma(w)) = \mathbf{G}(u_\gamma(w), v_\gamma(w)), \qquad \text{for all } w \in [0,1]. \tag{2.18}$$

Equation (2.18) shows that when two surface patches meet positionally along a boundary curve, there is a natural correspondence established between points in D_F and points in D_G. Namely, the point $(s_\gamma(w), t_\gamma(w))$ on γ_F is associated with the point $(u_\gamma(w), v_\gamma(w))$ on γ_G. This association defines a *correspondence map* M that carries points on γ_G into points on γ_F, written symbolically as

$$M : \gamma_F \rightarrow \gamma_G. \tag{2.19}$$

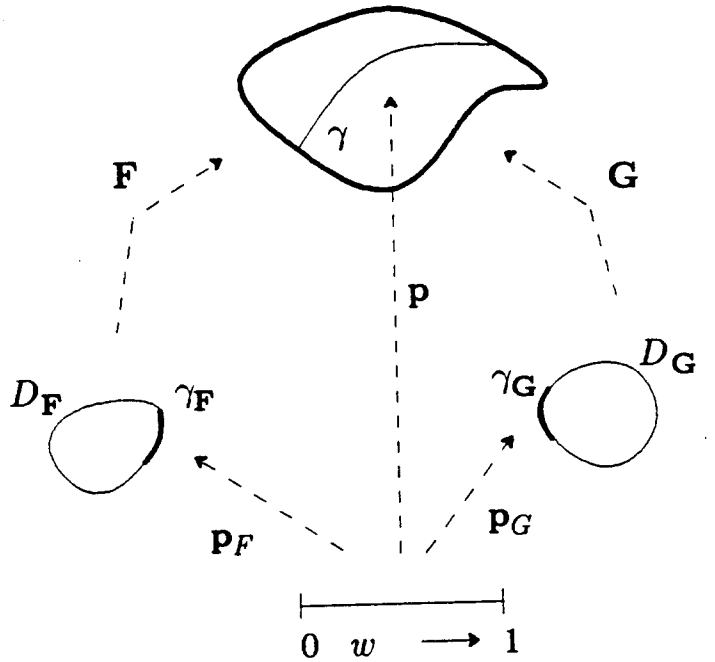

Figure 2.13. *γ has preimages γ_F and γ_G in D_F and D_G, respectively; γ_F and γ_G are generated parametrizations $\mathbf{p}_F(w)$ and $\mathbf{p}_G(w)$.*

Equation (2.19) shows that something interesting has occurred. The parameter spaces E_F and E_G were initially unrelated, but the fact that **F** and **G** meet with C^0 continuity along γ induces a partial relationship between E_F and E_G. As we will see shortly, knowing that **F** and **G** meet with parametric continuity along γ will tell us a great deal about the form of the correspondence map M. However, we must first define what it means for two patches to meet with parametric continuity.

Definition 2.2: *Let* $\mathbf{F}$ *and* $\mathbf{G}$ *be regular* C^∞ *parametrizations meeting with positional continuity on a boundary curve* γ *generated by a* C^∞ *parametrization* $\mathbf{p}(w)$, $w \in [0,1]$. $\mathbf{F}$ *and* $\mathbf{G}$ *are said to meet with* C^n *continuity at a point* $\mathbf{p}(w)$, $w \in [0,1]$, $n \geq 1$, *if*

$$\mathbf{F}^{(i,j)}(s_\gamma(w), t_\gamma(w)) = \mathbf{G}^{(i,j)}(u_\gamma(w), v_\gamma(w)), \qquad i+j = 1, ..., n. \tag{2.20}$$

They meet with C^n *continuity on* γ *if they meet with* C^n *continuity at* $\mathbf{p}(w)$ *for all* $w \in [0,1]$.

Since positional continuity (C^0) has been explicitly assumed, the $i+j=0$ case does not appear in equation (2.20). Also, we have only defined parametric continuity for patches that meet positionally along a differentiable curve. If the boundary curve is only piecewise differentiable, as shown in Figure 2.14, it may be treated on a piecewise basis using Definition 2.2.

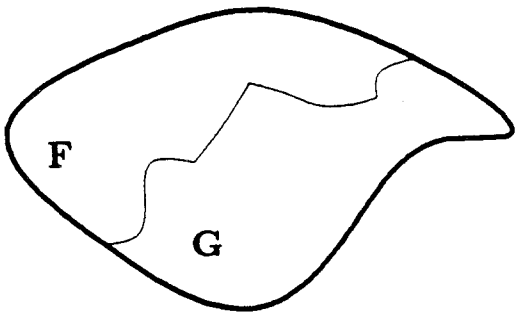

Figure 2.14. *The two patches above meet along a boundary curve that is piecewise* C^∞.

From equation (2.20) we can extract the relationship between γ_F and γ_G by determining the form of the correspondence map M. To do this, assume that $\mathbf{F}$ and $\mathbf{G}$ are as in Definition 2.2, meeting with at least C^1 continuity along γ. Equation (2.18) is a statement of equality for two differentiable functions of w, which implies that their derivatives are also equal. That is,

$$\frac{d\mathbf{F}}{dw} = \frac{d\mathbf{G}}{dw} \tag{2.21}$$

which can be expanded using the chain rule to yield

$$\frac{ds_\gamma}{dw}\frac{\partial \mathbf{F}}{\partial s} + \frac{dt_\gamma}{dw}\frac{\partial \mathbf{F}}{\partial t} = \frac{du_\gamma}{dw}\frac{\partial \mathbf{G}}{\partial u} + \frac{dv_\gamma}{dw}\frac{\partial \mathbf{G}}{\partial v}. \tag{2.22}$$

Since $\mathbf{F}$ and $\mathbf{G}$ are assumed to meet C^1 along γ, equation (2.22) can be written entirely in terms of derivatives of $\mathbf{F}$ as

$$\left(\frac{ds_\gamma}{dw} - \frac{du_\gamma}{dw}\right)\frac{\partial \mathbf{F}}{\partial s} + \left(\frac{dt_\gamma}{dw} - \frac{dv_\gamma}{dw}\right)\frac{\partial \mathbf{F}}{\partial t} = 0. \tag{2.23}$$

Regularity of $\mathbf{F}$ implies linear independence of its partial derivatives, so equation (2.23) holds if and only if

$$\begin{aligned} \frac{du_\gamma}{dw} &= \frac{ds_\gamma}{dw} \\ \frac{dv_\gamma}{dw} &= \frac{dt_\gamma}{dw}. \end{aligned} \tag{2.24}$$

Once again, equations (2.24) state equality of functions, so they may be integrated to give

$$\begin{aligned} u_\gamma(w) &= s_\gamma(w) + c_1 \\ v_\gamma(w) &= t_\gamma(w) + c_2 \end{aligned} \tag{2.25}$$

where c_1 and c_2 are constants of integration. Thus, if two parametrizations $\mathbf{F}$ and $\mathbf{G}$ meet with at least C^1 continuity along a boundary curve, then there exist constants c_1 and c_2 such that the equation for boundary correspondence (equation (2.18)) can be written as

$$\mathbf{F}(s_\gamma(w), t_\gamma(w)) = \mathbf{G}(s_\gamma(w) + c_1, t_\gamma(w) + c_2), \qquad \text{for all } w \in [0,1]. \tag{2.26}$$

If we let $\mathbf{q}$ be a point on γ_F, then from equation (2.25), the corresponding point $M(\mathbf{q})$ on γ_G is given by

$$M(\mathbf{q}) = \mathbf{q} + \mathbf{t}, \qquad \mathbf{q} \in \gamma_F, \tag{2.27}$$

where $\mathbf{t} = (c_1, c_2)$. Equation (2.27) states that the preimage boundary curves γ_F and γ_G can be brought into correspondence with a translation, implying that γ_F and γ_G have the *same shape*. By extending the domain of M to include all of E_F, we obtain an *inclusion map* $T : E_F \to E_G$ defined by

$$T(\mathbf{q}) = \mathbf{q} + \mathbf{t}, \qquad \mathbf{q} \in E_F. \tag{2.28}$$

The inclusion map T is a first order characterization of the correspondence between the domains D_F and D_G. Moreover, T is consistent with the induced correspondence map M determined from the knowledge that $\mathbf{F}$ and $\mathbf{G}$ meet with

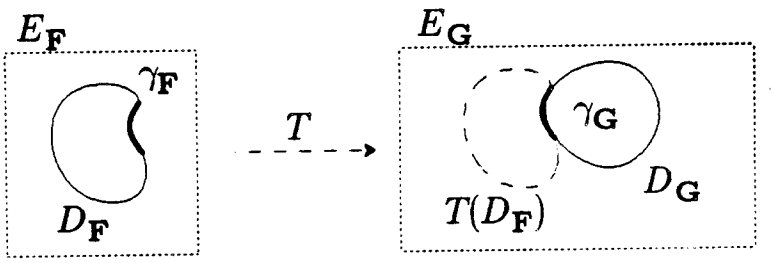

Figure 2.15. *The inclusion map T carries D_F into E_G by translation. The image of D_F under T is denoted by $T(D_F)$.*

C^1 continuity. T operates by adding a constant vector to the coordinates of a point in E_F to obtain the coordinates of the image point in E_G, as shown in Figure 2.15.

We take this opportunity to remark that there is a curious situation that can arise when stitching curve segments or surface patches together with parametric continuity that is well known in differential geometry, but does not seem to have been pointed out in the CAGD literature. It is possible for two curve segments or surface patches to meet with parametric continuity at the common joint or boundary curve without having smoothness of the *composite* curve or surface. Such a situation is shown in Figure 2.16. This type of pathology can easily be avoided for curves by requiring that the "ending point" $\mathbf{l}(u_1)$ meet the "starting point" $\mathbf{r}(t_0)$. In fact, this was the method used in Section 2.4. However, for surfaces the notions of starting point and ending point are inappropriate, so an alternate method must be chosen to avoid the pathological case of Figure 2.16. The method we adopt is based on the use of the inclusion map T from equation (2.28). It is relatively easy to show that the pathological case is impossible if $T(D_F)$ and D_G intersect only along γ_G, as shown in Figure 2.15. Stated intuitively, the pathological case is impossible if the parametric domains, when placed in correspondence, *abut* without overlapping. Note that this condition is violated in Figure 2.16. In practice, spline surfaces are always constructed so that domains abut without overlap, but to be completely correct, if a smooth composite surface is to be constructed, it is not sufficient to require only that the partial derivatives match up to a given order. Throughout the remainder of this chapter, we assume that domains abut when placed in correspondence by the inclusion map. The formalism of Chapter 6 rejects non-abutting domains in a natural way, thereby avoiding the pathological case of Figure 2.16.

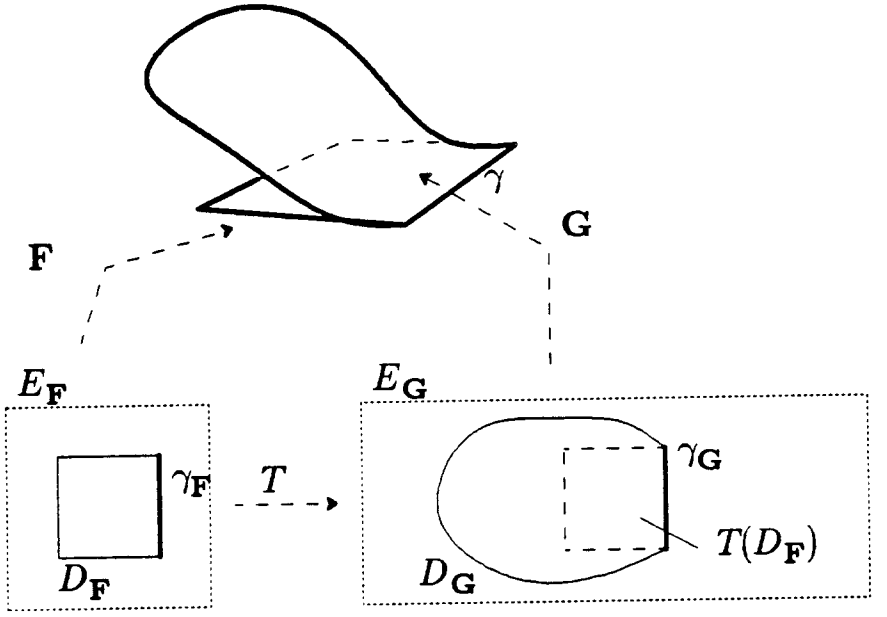

Figure 2.16. **F** *and* **G** *meet along a boundary curve* γ*, and partial derivatives up to order* 1 *agree all along* γ*. However, the inclusion map* T *(a translation) causes* $T(D_F)$ *and* D_G *to overlap, leading to a composite surface that is not smooth.*

2.5.2. Reparametrisation of Surface Patches

In anticipation of the definition of geometric continuity between surface patches (to be given in Section 2.5.3), we examine the bivariate reparametrization process. Two bivariate parametrizations are GO-equivalent if they have the same geometry and orientation in a neighborhood of each point on the surface patch. If $\mathbf{G} : D_G \subset E_G \rightarrow E$, having parameters (u, v), and $\tilde{\mathbf{G}} : D_{\tilde{G}} \subset E_{\tilde{G}} \rightarrow E$, having parameters $(\tilde{u}, \tilde{v})$ are GO-equivalent, then by the Inverse Function Theorem, they are related by

$$\tilde{\mathbf{G}}(\tilde{u}, \tilde{v}) = \mathbf{G}(u(\tilde{u}, \tilde{v}), v(\tilde{u}, \tilde{v})) \tag{2.29}$$

where the functions u and v, called the *change of variables* or the *change of parametrization,* describe a map carrying points in $D_{\tilde{G}}$ into points in D_G, as shown in Figure 2.17. The change of variables must satisfy the orientation preserving

condition **

$$u^{(1,0)}v^{(0,1)} - u^{(0,1)}v^{(1,0)} > 0. \tag{2.30}$$

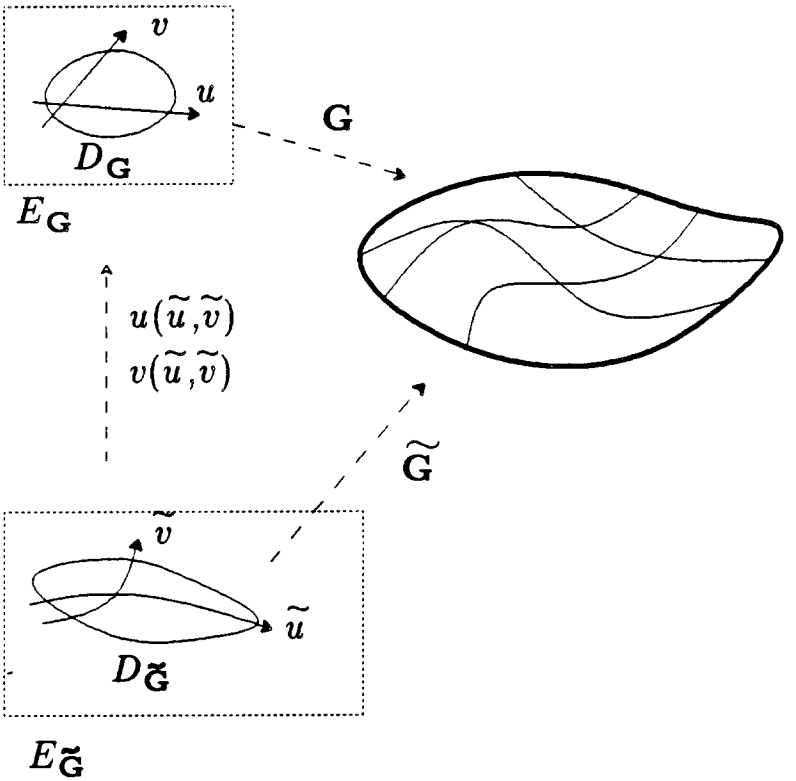

Figure 2.17. $\mathbf{G}$ *and* $\tilde{\mathbf{G}}$ *are GO-equivalent parametrizations related by the change of parametrization determined by* $u(\tilde{u},\tilde{v})$ *and* $v(\tilde{u},\tilde{v})$.

In complete analogy with curves, the bivariate chain rule can be used to express derivatives of $\tilde{\mathbf{G}}$ in terms of $\mathbf{G}$. For example, the first order partial derivatives are given by

$$\begin{aligned}\frac{\partial\tilde{\mathbf{G}}}{\partial\tilde{u}} &= \frac{\partial u}{\partial\tilde{u}}\frac{\partial\mathbf{G}}{\partial u} + \frac{\partial v}{\partial\tilde{u}}\frac{\partial\mathbf{G}}{\partial v}\\ \frac{\partial\tilde{\mathbf{G}}}{\partial\tilde{v}} &= \frac{\partial u}{\partial\tilde{v}}\frac{\partial\mathbf{G}}{\partial u} + \frac{\partial v}{\partial\tilde{v}}\frac{\partial\mathbf{G}}{\partial v}.\end{aligned} \tag{2.31}$$

In general, the i,j^{th} partial derivative of $\tilde{\mathbf{G}}$ can be expressed as some function, call it $\mathcal{CR}_{i,j}$, of the partial derivatives of $\mathbf{G}$, u, and v, up to order $i+j$. Stated

** Readers familiar with multivariate calculus may recognize equation (2.30) as the Jacobian of the change of variables (see Section 6.3).

mathematically,

$$\widetilde{\mathbf{G}}^{(i,j)} = \mathcal{CR}_{i,j}(\mathbf{G}^{(k,l)}, u^{(k,l)}, v^{(k,l)}), \tag{2.32}$$

where the indices (k,l) are to take on all positive values such that $k+l = i+j$.

2.5.3. Geometric Continuity for Surface Patches

Just as for curves, parametric continuity is appropriate for the GOR model of a surface, but it is not suitable for use with the GO model since it places emphasis on irrelevant rate information. The determination of continuity can be made insensitive to rate by allowing reparametrization. More formally:

Definition 2.3: *Let* $\mathbf{F}$ *and* $\mathbf{G}$ *be regular* C^∞ *parametrizations meeting along a boundary curve* γ*, such that* γ *can be generated by a regular* C^∞ *parametrization* $\mathbf{p}(w)$*.* $\mathbf{F}$ *and* $\mathbf{G}$ *are said to meet with* G^n *continuity at a point* $\mathbf{p}(w)$ *if and only if there exist GO-equivalent parametrizations* $\widetilde{\mathbf{F}}$ *and* $\widetilde{\mathbf{G}}$ *that meet with* C^n *continuity at* $\mathbf{p}(w)$*. They meet with* G^n *continuity on* γ *if there exist GO-equivalent parametrizations that meet with* C^n *continuity on* γ*.*

Remark 2.1: *There is another reasonable definition of geometric continuity along* γ*:* $\mathbf{F}$ *and* $\mathbf{G}$ *meet with* G^n *continuity on* γ *if they meet with* G^n *continuity at every point* $\mathbf{p}(w)$ *of* γ*. Note that this definition allows the GO-equivalent parametrizations used at point a* $\mathbf{p}_1$ *to differ from the GO-equivalent parametrizations used at point a* $\mathbf{p}_2$*. It is conjectured that this definition is equivalent to Definition 2.3, but a proof has not yet been constructed.*

In complete analogy with curves, only one of the parametrizations actually needs to be reparametrized. This implies that $\mathbf{F}$ and $\mathbf{G}$ meet with G^n continuity on γ if and only if there exists a $\widetilde{\mathbf{G}}$ such that

$$\mathbf{F}^{(i,j)}(s_\gamma(w), t_\gamma(w)) = \widetilde{\mathbf{G}}^{(i,j)}(\widetilde{u}_\gamma(w), \widetilde{v}_\gamma(w)), \qquad i+j = 1, ..., n \tag{2.33}$$

for all $w \in [0,1]$, where $(\widetilde{u}_\gamma(w), \widetilde{v}_\gamma(w))$ denotes the preimage of $\mathbf{p}(w)$ in $\widetilde{\mathbf{G}}$'s domain. To emphasize the relationship between the preimage curves established in equation (2.26), equation (2.33) is better written as

$$\mathbf{F}^{(i,j)}(s_\gamma(w), t_\gamma(w)) = \widetilde{\mathbf{G}}^{(i,j)}(s_\gamma(w) + c_1, t_\gamma(w) + c_2), \tag{2.34}$$

$$i+j = 1, ..., n, \qquad w \in [0,1].$$

To simplify the following discussion, let $\mathbf{F}(s,t)$ be such that the boundary curve γ is generated by holding s fixed at s_1, letting t vary to trace out the curve. That is, γ is generated by $\mathbf{p}(t) = \mathbf{F}(s_1,t)$, $t \in [t_0,t_1]$. The general situation will be addressed at the end of this section. With the above restriction, equation (2.18) then becomes

$$\mathbf{F}(s_1,t) = \mathbf{G}(u_\gamma(t), v_\gamma(t)), \qquad t \in [t_0,t_1] \tag{2.35}$$

and equation (2.34) becomes

$$\mathbf{F}^{(i,j)}(s_1,t) = \tilde{\mathbf{G}}^{(i,j)}(s_1+c_1, t+c_2), \qquad i+j = 1,...,n, \quad t \in [t_0,t_1]. \tag{2.36}$$

The annoying constants c_1 and c_2 will vanish when we differentiate, so their value is immaterial. For convenience, we choose $c_1 = -s_1 + \tilde{u}_0$ and $c_2 = 0$ to get

$$\mathbf{F}^{(i,j)}(s_1,t) = \tilde{\mathbf{G}}^{(i,j)}(\tilde{u}_0,t), \qquad i+j = 1,...,n, \quad t \in [t_0,t_1], \tag{2.37}$$

for some new constant $\tilde{u}_0$. Of course, the value of $\tilde{u}_0$ is also irrelevant; it was chosen to point out symmetry in equation (2.37).

Equation (2.36) implies that there are many conditions that must be satisfied if $\mathbf{F}$ and $\tilde{\mathbf{G}}$ are to meet with C^n continuity. However, many of the conditions are consequences of others. In particular, if we know that $\mathbf{F}^{(i,0)}(s_1,t) = \tilde{\mathbf{G}}^{(i,0)}(\tilde{u}_0,t)$ for some i, then by differentiating j times with respect to t, we know that $\mathbf{F}^{(i,j)}(s_1,t) = \tilde{\mathbf{G}}^{(i,j)}(\tilde{u}_0,t)$. In other words, $\mathbf{F}$ and $\tilde{\mathbf{G}}$ meet with C^n continuity on γ if and only if

$$\mathbf{F}^{(i,0)}(s_1,t) = \tilde{\mathbf{G}}^{(i,0)}(\tilde{u}_0,t), \qquad i = 1,...,n, \quad t \in [t_0,t_1]. \tag{2.38}$$

Equation (2.38) can be rewritten in terms of derivatives of $\mathbf{G}$ by using the chain rule expansion from equation (2.32):

$$\begin{aligned} \mathbf{F}^{(i,0)}(s_1,t) &= \mathcal{CR}_{i,0}(\mathbf{G}^{(k,l)}(u_\gamma(t), v_\gamma(t),), u^{(k,l)}(\tilde{u}_0,t), v^{(k,l)}(\tilde{u}_0,t)), \\ &\quad i = 1,...,n, \\ &\quad t \in [t_0,t_1]. \end{aligned} \tag{2.39}$$

It is easily shown that the only derivatives of u and v that appear in equation (2.39) are $u^{(k,0)}(\tilde{u}_0,t)$ and $v^{(k,0)}(\tilde{u}_0,t)$. If we let

$$\begin{aligned} \beta_{u,k}(t) &= u^{(k,0)}(\tilde{u}_0,t) \\ \beta_{v,k}(t) &= v^{(k,0)}(\tilde{u}_0,t), \end{aligned} \tag{2.40}$$

and note that

$$\begin{aligned} \beta_{u,0}(t) &\equiv u(\widetilde{u}_0, t) = u_\gamma(t) \\ \beta_{v,0}(t) &\equiv v(\widetilde{u}_0, t) = v_\gamma(t), \end{aligned} \tag{2.41}$$

then equation (2.39) can be written as

$$\begin{aligned} \mathbf{F}^{(i,0)}(s_1, t) = \mathcal{CR}_{i,0}(\mathbf{G}^{(k,l)}(\beta_{u,0}(t), \beta_{v,0}(t)), \beta_{u,k}(t), \beta_{v,k}(t)), \\ i = 1, ..., n, \\ t \in [t_0, t_1]. \end{aligned} \tag{2.42}$$

With this notation, the orientation preserving quality of the change of variables (from equation (2.30)) becomes

$$\beta_{u,1}(t)\,\beta^{(1)}_{v,0}(t) - \beta^{(1)}_{u,0}(t)\,\beta_{v,1}(t) > 0, \tag{2.43}$$

where derivatives of the β's refer to derivatives with respect to t.

The conditions implied by equation (2.42) are the *bivariate Beta constraints.* More formally:

Theorem 2.2: *Let* $\mathbf{F}$ *and* $\mathbf{G}$ *be as in Definition 2.3.* $\mathbf{F}$ *and* $\mathbf{G}$ *meet with* G^n *continuity on* γ *if and only if there exist* C^∞ *functions* $\beta_{u,i}(t)\,, \beta_{v,i}(t)$, $i = 1, ..., n$ *such that equations (2.42) are satisfied, subject to equation (2.43).*

We have only argued necessity here; a detailed, complete proof is deferred to Chapter 6.

Just as for curves, the β's are the shape parameters, with the important difference that for surfaces the shape parameters are actually functions defined all along the boundary curve. Thus, when stitching two surface patches together with G^n continuity, $2n$ shape parameters (functions) are introduced.

Remark 2.2: *Theorem 2.2 implies that the* β*'s with index larger than 0 can be arbitrarily chosen functions. However, the* β*'s with 0 index are uniquely determined by equation (2.41), hence they are fixed by the assumption of* C^0 *continuity. In essence, the approach we have taken initially assumes* C^0 *continuity, then imposes restrictions on the parametrizations to achieve continuity of higher order. As mentioned, the assumption of* C^0 *continuity fixes* $\beta_{u,0}(t)$ *and* $\beta_{v,0}(t)$*, by fixing the correspondence map that carries points on* γ_F *into points on* γ_G*.*

Perhaps it is more in keeping with the spirit of geometric continuity not to assume C^0 *continuity initially, requiring only that there exist functions* $\beta_{u,0}(t)$ *and*

$\beta_{v,0}(t)$ such that **F** *and* **G** *meet with positional continuity. One of the attractive aspects of such an approach is that the β_0's enter as arbitrary parameters, just as the β's with higher indices. At first, it appears that such an approach introduces more shape parameters than the approach we have adopted. However, given such a scheme, and given a particular pair of parametrizations* **F** *and* **G**, *the first step in specifying their continuity properties is to choose the β_0's that make them meet with positional continuity, thereby fixing the β_0's. The β_0's are then used in the higher order constraints. We have simply chosen to assume the first step has already been done, and that the β_0's are given.*

The distinction between the two schemes makes little difference when only two parametrizations are involved. However, the assumption of C^0 continuity greatly simplifies the formal development of Chapter 6 wherein an entire collection of parametrizations are dealt with.

In Section 2.4, some simple heuristics were given for determining the i^{th} univariate Beta constraint by a peculiar kind of differentiation of the $i-1^{\text{st}}$ constraint. A similar set of heuristics can be obtained for the bivariate case. In particular, the i^{th} constraint can be obtained from the $i-1^{\text{st}}$ constraint by "differentiating" with respect to $\tilde{u}$, using the following heuristic rules

$$\begin{aligned}
\frac{\partial \mathbf{F}^{(i,0)}}{\partial \tilde{u}} &= \mathbf{F}^{(i+1,0)} \\
\frac{\partial \mathbf{G}^{(i,j)}}{\partial \tilde{u}} &= \beta_{u,1}\, \mathbf{G}^{(i+1,j)} + \beta_{v,1}\, \mathbf{G}^{(i,j+1)} \\
\frac{\partial \beta^k_{u,1}}{\partial \tilde{u}} &= k\, \beta^{k-1}_{u,i}\, \beta_{u,i+1} \\
\frac{\partial \beta^k_{v,1}}{\partial \tilde{u}} &= k\, \beta^{k-1}_{v,i}\, \beta_{v,i+1}.
\end{aligned} \tag{2.44}$$

The first rule simply states that the chain rule is not to be used on the left side of the constraint. The next rule is a restatement of the chain rule for derivatives of **G**. The last two rules reflect the fact that higher order shape parameters result from higher order derivatives of the change of variables with respect to the cross boundary variable.

The above derivation assumed that the boundary curve corresponded to a parametric direction of the parametrization **F**. If **F**'s boundary curve does not correspond to a parametric direction, it is always possible to find a GO-equivalent parametrization $\widetilde{\mathbf{F}}$ whose boundary curve does. The Beta constraints above can

be used to relate derivatives of $\widetilde{\mathbf{F}}$ and $\mathbf{G}$. The constraints can then be restated in terms of derivatives of $\mathbf{F}$ using the inverse reparametrization and the chain rule. Although this can be done in principle, it may be computationally prohibitive for high order continuity. This does not seem to be damaging in practice since all currently implemented techniques (that we know of) assume that the boundary curves of both patches correspond to parametric directions.

2.5.3.1. Equivalence with Previous Measures

In this section, we sketch a proof showing that our definitions of geometric continuity reduce to the previous definitions of tangent plane and osculating paraboloid continuity. Actually, our definitions are equivalent to continuity of *oriented* tangent plane and osculating paraboloid continuity. Continuity of oriented tangent planes is slightly stronger than continuity of (unoriented) tangent planes in that continuity of tangent planes is equivalent to requiring that the unit normals either align or anti-align, while continuity of oriented tangent planes only allows alignment of the unit normals. The reader is referred to Section 6.4.1 for a more complete discussion of orientation.

We begin by assuming that $\mathbf{F}$ and $\mathbf{G}$ meet with G^1 continuity on γ, implying that there exist $\widetilde{\mathbf{F}}$ and $\widetilde{\mathbf{G}}$ that meet with C^1 continuity on γ. Since the first order partial derivatives of $\widetilde{\mathbf{F}}$ and $\widetilde{\mathbf{G}}$ agree at every point on γ, equation (2.3) implies that $\widetilde{\mathbf{F}}$ and $\widetilde{\mathbf{G}}$ have a common unit normal vector at every point. The unit normal is invariant under GO-equivalent reparametrization, implying that $\mathbf{F}$ and $\mathbf{G}$ have a common unit normal at every point. This argument shows that G^1 continuity is sufficient for unit normal continuity, or equivalently, oriented tangent plane continuity.

To show necessity, assume that $\mathbf{F}$ and $\mathbf{G}$ have a common unit normal at each point. The first step is to choose $\widetilde{\mathbf{F}}$, GO-equivalent to $\mathbf{F}$, such that γ is generated by $\widetilde{\mathbf{F}}(0,\tilde{t})$. The fact that $\widetilde{\mathbf{F}}$ and $\mathbf{G}$ have a common unit normal implies that there exist functions $\alpha_1(\tilde{t})$, $\alpha_2(\tilde{t})$, $\alpha_3(\tilde{t})$, and $\alpha_4(\tilde{t})$, such that

$$\begin{aligned}\widetilde{\mathbf{F}}^{(1,0)}(0,\tilde{t}) &= \alpha_1(\tilde{t})\,\mathbf{G}^{(1,0)}(u_\gamma(\tilde{t}),v_\gamma(\tilde{t})) + \alpha_2(\tilde{t})\,\mathbf{G}^{(0,1)}(u_\gamma(\tilde{t}),v_\gamma(\tilde{t}))\\ \widetilde{\mathbf{F}}^{(0,1)}(0,\tilde{t}) &= \alpha_3(\tilde{t})\,\mathbf{G}^{(1,0)}(u_\gamma(\tilde{t}),v_\gamma(\tilde{t})) + \alpha_4(\tilde{t})\,\mathbf{G}^{(0,1)}(u_\gamma(\tilde{t}),v_\gamma(\tilde{t}))\end{aligned} \tag{2.45}$$

and

$$\alpha_1(\tilde{t})\,\alpha_4(\tilde{t}) - \alpha_2(\tilde{t})\,\alpha_3(\tilde{t}) > 0 \tag{2.46}$$

for all $t \in [t_0,t_1]$. The reasoning is as follows. The partial derivatives $\mathbf{G}^{(1,0)}$ and $\mathbf{G}^{(0,1)}$ define the tangent plane of $\mathbf{G}$ at $\mathbf{p}(\tilde{t})$. Since $\mathbf{G}$ is assumed to be regular,

these vectors are linearly independent; hence, they span the tangent plane. If $\widetilde{\mathbf{F}}$ is to share this tangent plane, its first order partial derivatives must be expressible as linear combinations of $\mathbf{G}^{(1,0)}$ and $\mathbf{G}^{(0,1)}$, as shown in equation (2.45). The restriction (2.46) on the α's is necessary to ensure that the normal vectors of $\widetilde{\mathbf{F}}$ and $\mathbf{G}$ align rather than anti-align.

The second equation of (2.45) is a consequence of C^0 continuity; it follows from differentiation of

$$\widetilde{\mathbf{F}}(0,\widetilde{t}) = \mathbf{G}(\beta_{u,0}(\widetilde{t}),\beta_{v,0}(\widetilde{t})) \tag{2.47}$$

with respect to $\widetilde{t}$, yielding

$$\widetilde{\mathbf{F}}^{(0,1)}(0,\widetilde{t}) = \beta_{u,0}(\widetilde{t})\ \mathbf{G}^{(1,0)}(\beta_{u,0}(\widetilde{t}),\beta_{v,0}(\widetilde{t})) + \beta_{v,0}(\widetilde{t})\ \mathbf{G}^{(0,1)}(\beta_{u,0}(\widetilde{t}),\beta_{v,0}(\widetilde{t})). \tag{2.48}$$

where $(\beta_{u,0}(\widetilde{t}),\beta_{v,0}(\widetilde{t}))$ is a parametrization for γ_G, parametrized by $\widetilde{t}$. Comparing (2.48) to the first equation of (2.45), and using the linear independence of $\mathbf{G}^{(1,0)}$ and $\mathbf{G}^{(0,1)}$, it must be that $\alpha_3(\widetilde{t}) = \beta_{u,0}(\widetilde{t})$ and $\alpha_4 = \beta_{v,0}(\widetilde{t})$. The first equation of (2.45) is then seen to be the first order Beta constraint generated by Rules (2.44), where $\alpha_1 = \beta_{u,1}$ and $\alpha_2 = \beta_{v,1}$. Theorem 2.2 then guarantees that $\widetilde{\mathbf{F}}$ and $\mathbf{G}$ meet with G^1 continuity, and since $\widetilde{\mathbf{F}}$ and $\mathbf{F}$ are GO-equivalent, $\mathbf{F}$ and $\mathbf{G}$ must also meet with G^1 continuity, thus completing the proof of first order equivalence.

Technical Note: *Actually, before Theorem 2.2 can be invoked, we must verify that α_1 and α_2 are C^∞ functions. This can be done by letting $\mathbf{v}(\widetilde{t})$ be a C^∞ vector function that is perpendicular to $\mathbf{G}^{(1,0)}(\beta_{u,0}(\widetilde{t}),\beta_{v,0}(\widetilde{t}))$, but not perpendicular to $\mathbf{G}^{(0,1)}(\beta_{u,0}(\widetilde{t}),\beta_{v,0}(\widetilde{t}))$. Such a $\mathbf{v}(\widetilde{t})$ must exist because $\mathbf{G}^{(1,0)}$ is a C^∞ function, and the first order partial derivatives of $\mathbf{G}$ are linearly independent. By dotting the first equation of (2.45) with $\mathbf{v}(\widetilde{t})$, the term containing $\mathbf{G}^{(1,0)}$ vanishes, leaving (after rearrangement)*

$$\alpha_2(\widetilde{t}) = \frac{\widetilde{\mathbf{F}}^{(1,0)}(0,\widetilde{t}) \cdot \mathbf{v}(\widetilde{t})}{\mathbf{G}^{(0,1)}(u_\gamma(\widetilde{t}),v_\gamma(\widetilde{t})) \cdot \mathbf{v}(\widetilde{t})}. \tag{2.49}$$

The right side of equation (2.49) is a C^∞ function, implying that $\alpha_2(\widetilde{t})$ is C^∞. In a similar way, $\alpha_1(\widetilde{t})$ can be shown to be a C^∞ function.

To establish second order equivalence, we begin by assuming that $\mathbf{F}$ and $\mathbf{G}$ meet with G^2 continuity on γ; we must show that they have common unit normal and osculating paraboloids at each point along the boundary. Continuity of unit normals was established above, so we must simply show that $\mathbf{F}$ and $\mathbf{G}$ have a

common osculating paraboloid at each point. This follows from the fact that there exist $\widetilde{\mathbf{F}}$ and $\widetilde{\mathbf{G}}$ that meet with C^2 continuity along the boundary. From equation (2.5), $\widetilde{\mathbf{F}}$ and $\widetilde{\mathbf{G}}$ must have a common osculating paraboloid at each point on the boundary curve, and since the osculating paraboloid is invariant under GO-equivalent reparametrization, $\mathbf{F}$ and $\mathbf{G}$ must have a common osculating paraboloid at each point.

Proving the converse is slightly more complicated. We begin with the assumption that $\mathbf{F}$ and $\mathbf{G}$ have a common unit normal and osculating paraboloid at each point on the boundary. We must show that this assumption guarantees the existence of GO-equivalent parametrizations that meet with C^2 continuity on γ. Recall from Section 2.2 that the osculating paraboloid for $\mathbf{F}$ is conveniently expressed in the coordinate system $(\mathbf{F}^{(1,0)}, \mathbf{F}^{(0,1)}, \widehat{\mathbf{N}})$. If we reparametrize $\mathbf{G}$ to obtain $\widetilde{\mathbf{G}}$ that meets $\mathbf{F}$ with C^1 continuity, then the coordinate system $(\widetilde{\mathbf{G}}^{(1,0)}, \widetilde{\mathbf{G}}^{(0,1)}, \widehat{\mathbf{N}})$ is identical to the coordinate system for $\mathbf{F}$. Relative to this coordinate system, we can equate the osculating paraboloids of $\mathbf{F}$ and $\widetilde{\mathbf{G}}$ using equation (2.4) to obtain

$$L_F x^2 + 2M_F xy + N_F y^2 = L_{\widetilde{\mathbf{G}}} x^2 + 2M_{\widetilde{\mathbf{G}}} xy + N_{\widetilde{\mathbf{G}}} y^2. \tag{2.50}$$

Equation (2.50) states equality of polynomials, so the coefficients must be equal. That is, $L_F = L_{\widetilde{\mathbf{G}}}$, $M_F = M_{\widetilde{\mathbf{G}}}$, and $N_F = N_{\widetilde{\mathbf{G}}}$. Equality of the M's and N's can be shown to follow from the fact that $\mathbf{F}$ and $\widetilde{\mathbf{G}}$ meet with C^0 continuity along γ. The only new information is equality of the L's, which when written out becomes

$$\mathbf{F}^{(2,0)} \cdot \widehat{\mathbf{N}} = \widetilde{\mathbf{G}}^{(2,0)} \cdot \widehat{\mathbf{N}}. \tag{2.51}$$

Thus, $\mathbf{F}^{(2,0)}$ and $\widetilde{\mathbf{G}}^{(2,0)}$ can only differ by a component perpendicular to $\widehat{\mathbf{N}}$, or equivalently, by a component in the tangent plane. Since the first order partial derivatives of $\widetilde{\mathbf{G}}$ span the tangent plane, any vector in the tangent plane can be expressed as a linear combination of them. The second order partial derivatives in equation (2.51) must be therefore related by

$$\mathbf{F}^{(2,0)} = \widetilde{\mathbf{G}}^{(2,0)} + \alpha_1 \widetilde{\mathbf{G}}^{(1,0)} + \alpha_2 \widetilde{\mathbf{G}}^{(0,1)} \tag{2.52}$$

for some functions α_1 and α_2 defined along the boundary curve. The functions α_1 and α_2 can be shown to be C^∞ using the trick above of dotting with a vector $\mathbf{v}(t)$. Equation (2.52) is then identified as a special case of the second order Beta constraint where $\beta_{u,1} = 1$, $\beta_{v,1} = 0$, $\beta_{u,2} = \alpha_1$, and $\beta_{v,2} = \alpha_2$, implying that

there exists a $\tilde{\tilde{\mathbf{G}}}$ (GO-equivalent to $\tilde{\mathbf{G}}$) that meets $\mathbf{F}$ with C^2 continuity on γ. $\tilde{\tilde{\mathbf{G}}}$ is also GO-equivalent to $\mathbf{G}$, and therefore $\mathbf{F}$ and $\mathbf{G}$ meet with G^2 continuity on γ.

To reiterate the results of this section, G^1 continuity is equivalent to requiring common unit normal vectors, and G^2 continuity is equivalent to requiring common unit normals and common osculating paraboloids, or equivalently, common Dupin indicatrices. Thus, the chain rule approach agrees with geometric intuition for both G^1 and G^2 continuity. For higher order continuity, geometric intuition becomes more feeble, but the chain rule still applies.

2.6. Summary

We have defined n^{th} order geometric continuity for parametric curves and surfaces, and derived the Beta constraints that are necessary and sufficient for it. The derivation of the Beta constraints is based on the simple principle of reparametrization in conjunction with the univariate chain rule for curves, and the bivariate chain rule for surfaces. This approach therefore uncovers the connection between geometric continuity for curves and geometric continuity for surfaces. The approach also provides new insight into the nature of geometric continuity in general, and allows the determination of the Beta constraints with less effort than previously required.

The use of the Beta constraints for G^n continuity allows the introduction of n shape parameters for curves, and $2n$ shape parameters for surfaces. Intuitively, the shape parameters determine the cross boundary relationship between the parameter lines on the respective curve segments or surface patches.

The shape parameters can be set arbitrarily. Therefore, they may be used to modify the shape of a geometrically continuous curve or surface. However, geometric continuity is only appropriate for applications where the "rate" of a parametrization is unimportant since discontinuities in rate are allowed. Examples of the use of geometric continuity and the Beta constraints for curve and surface techniques are given in Chapters 3, 4, and 5.

The approach we have taken is not based on measures that are inherent to curves and surfaces, so the generalization to p-variate objects (volumes, hypervolumes, etc.) can be made very simply: two p-variate parametrizations are GO-equivalent if and only if they are related by a change of parametrization with positive Jacobian. The corresponding Beta constraints may be derived in complete analogy to the development of Section 2.5, using the p-variate chain rule [16] in

place of the bivariate chain rule. This is one of the topics addressed in Chapter 6.

3

Spline Curves

In this chapter, some specific uses of the univariate Beta constraints are examined. These include the placement of Bézier control vertices, the construction of the cubic Beta-spline basis segments, and a brief discussion of the class of geometrically continuous Catmull-Rom splines.

3.1. Background

We focus attention on spline curves that are formed as a weighted average of *control vertices.* These *blended splines* take the form

$$\mathbf{Q}(u) = \sum_{i=0}^{m} \mathbf{V}_i W_i(u), \qquad u \in [u_0, u_f], \tag{3.1}$$

where the control vertices $\mathbf{V}_i$, $i = 0, ..., m$ are chosen by the designer from $\Re^2$ or $\Re^3$. The sequence $< \mathbf{V}_i >_{i=0}^{m}$ is called a *control polygon* and the functions $W_i(u)$ are called *blending* or *basis* functions.

In this chapter we will deal exclusively with piecewise polynomial basis functions. The piecewise nature of the basis functions gives rise to a piecewise parametric function $\mathbf{Q}$. The segments of the basis functions are called *basis segments,* each of which is a (non-piecewise) polynomial. A typical basis function is plotted in Figure 3.1. Referring to Figure 3.1, the breakpoints between basis segments are called *knots*; the image of a knot is called a *joint.* The knots partition $[u_0, u_f]$ into smaller intervals $[u_j, u_{j+1}]$.

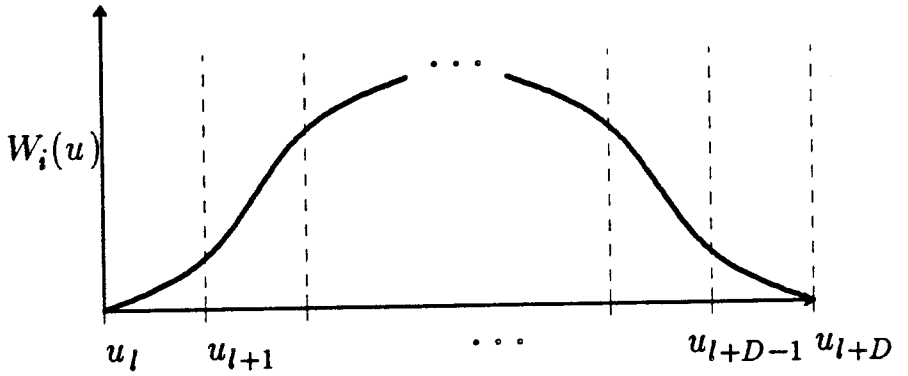

Figure 3.1. *Plotted above is a typical basis function $W_i(u)$. The knots, labeled $u_l, u_{l+1}, \ldots$, determine the points where basis segments meet.*

The blending functions naturally determine the character of the resulting spline. If the blending functions have *local support* (that is, they are nonzero only over a subrange of $[u_0, u_f]$), then a perturbation of a control vertex induces a local perturbation on $\mathbf{Q}$. This is known as the *property of local control.* To obtain a spline that is independent of the coordinate system in which the vertices are expressed, the basis functions must form a *partition of unity*; that is, they must satisfy

$$\sum_{j=0}^{m} W_j(u) = 1 \qquad \forall u \in [u_0, u_f]. \tag{3.2}$$

If the basis functions are non-negative and form a partition of unity, then the curve must lie in the *convex hull** * of the control polygon. This is referred to as the *convex hull property.* Finally, the spline may either be *interpolating* or *approximating.* Interpolating splines are guaranteed to pass through the vertices. Approximating splines generally do not interpolate all the control vertices. Rather, an approximating spline typically represents a "smoothed" version of its defining control polygon.

Not all splines are of the blended form given in equation (3.1). For instance, it is common to space the knots of a *cubic interpolatory spline* (cf. deBoor [13]) based on the distance between control vertices. The dependence of the spline on the control vertices is non-linear, so it cannot be written in the form of equation (3.1). It is also possible to construct non-polynomial splines such as the *spline*

* *The convex hull of a set of points is the smallest convex set containing the points.

under tension due to Schweikert [54].

Assuming that the r^{th} segment of a blended spline $\mathbf{Q}$, denoted by $\mathbf{q}_r$, is generated when the domain variable varies on $[u_r, u_{r+1}]$, we can write $\mathbf{q}_r$ in terms of the basis segments of $W_i(u)$ as

$$\mathbf{Q}(u) = \mathbf{q}_r(u) = \sum_{i=0}^{m} \mathbf{V}_i w_{i,k}(u), \qquad u \in [u_r, u_{r+1}] \tag{3.3}$$

where $w_{i,k}(u)$ is the segment of $W_i(u)$ that is supported on the interval $[u_r, u_{r+1}]$. Without loss of generality, we may parametrize each $w_{i,k}(u)$ on $[0,1]$, implying that $\mathbf{q}_r$ is parametrized on $[0,1]$. Assuming a $[0,1]$ parametrization for the basis segments, equation (3.3) becomes

$$\mathbf{q}_r(u) = \sum_{i=0}^{m} \mathbf{V}_i w_{i,k}(u), \qquad u \in [0,1]. \tag{3.4}$$

This *segment definition* of a spline is often more useful than the *piecewise definition* of equation (3.1).

A common special case of the blended spline occurs when $W_i(u)$ is a translated version of a canonical blending function $W(u)$; that is, $W_i(u) = W(u-i)$. A spline of this type is said to be *uniform*. In this case, equation (3.2) becomes

$$\sum_{i=0}^{m} W(u-i) = 1, \qquad u \in [u_0, u_f]. \tag{3.5}$$

If the basis segments $w_k(u)$ of $W(u)$ are parametrized on $[0,1]$, equation (3.5) becomes

$$\sum_{k} w_k(u) = 1, \qquad u \in [0,1] \tag{3.6}$$

where the sum is taken over all indices k such that $w_k(u)$ is a segment of W.

3.1.1. Bézier Curves

Bézier curves of degree d are defined as

$$\mathbf{q}(u) = \sum_{i=0}^{d} \mathbf{V}_i b_i^d(u), \qquad u \in [0,1]. \tag{3.7}$$

where the Bézier basis functions are the Bernstein polynomials of degree d given by

$$b_i^d(u) = \binom{d}{i} u^i (1-u)^{d-i}. \tag{3.8}$$

Lower case $\mathbf{q}$ and b have been used in equation (3.7) because, strictly speaking, Bézier curves are not piecewise. In Section 3.2 we discuss how separate Bézier curves can be strung together to produce a geometrically continuous spline.

Bézier curves have the following useful properties:

1) *Interpolation of end vertices:* $\mathbf{q}(0) = \mathbf{V}_0$ and $\mathbf{q}(1) = \mathbf{V}_d$.
2) *Derivatives:* The function $\mathbf{q}^{(1)}(u)$ is a parametric polynomial of degree $d-1$ given by

$$\mathbf{q}^{(1)}(u) = \sum_{i=0}^{d-1} \mathbf{V}_i^{[1]} b_i^{d-1}(u), \qquad u \in [0,1] \tag{3.9}$$

where $\mathbf{V}_i^{[1]} = d(\mathbf{V}_{i+1} - \mathbf{V}_i)$. By property 1) above,

$$\begin{aligned} \mathbf{q}^{(1)}(0) &= \mathbf{V}_0^{[1]} = d(\mathbf{V}_1 - \mathbf{V}_0) \\ \mathbf{q}^{(1)}(1) &= \mathbf{V}_{d-1}^{[1]} = d(\mathbf{V}_d - \mathbf{V}_{d-1}) \end{aligned} \tag{3.10}$$

Higher derivatives follow from repeated application of equation (3.9).

3.1.2. B-spline Curves

B-spline curves are blended splines where the B-spline basis functions of order k (order = degree + 1), denoted by $N_i^k(u)$, may be defined recursively by the *Cox/deBoor relation* [11,20]:

$$N_i^k(u) = \frac{u - u_i}{u_{i+k-1} - u_i} N_i^{k-1}(u) + \frac{u_{i+k} - u}{u_{i+k} - u_{i+1}} N_{i+1}^{k-1}(u) \quad \text{for } k \neq 1 \tag{3.11}$$

$$N_i^1(u) = \begin{cases} 1 & u_i \leq u < u_{i+1} \\ 0 & \text{otherwise.} \end{cases}$$

More precisely, given a control polygon $\mathbf{V}_0, \ldots, \mathbf{V}_m$, and an *extended knot vector* $\Delta = (u_{-k+1}, \ldots, u_{m+k-1})$, the B-spline curve of order k is given by

$$\mathbf{Q}(u) = \sum_{i=0}^{m} \mathbf{V}_i N_i^k(u), \qquad u \in [u_0, u_m]. \tag{3.12}$$

B-spline basis functions can be shown to form a partition of unity, are non-negative, and have local support. The B-spline curve is therefore coordinate system independent, lies in the convex hull of the control polygon, and has local control.

The continuity of the basis functions, and hence, continuity of the B-spline curve, is determined by the polynomial degree of the basis functions and by the knot vector. The *multiplicity* of a knot is the number of times the knot appears in the knot vector. The continuity of the basis functions of order k at a knot of multiplicity μ is $C^{k-\mu-1}$; the resulting B-spline curve inherits this continuity.

Example 3.1: *Uniform Cubic B-spline: The uniform cubic B-spline results when* $k = 4$ *and* $\Delta = (-3, -2, ..., m+2, m+3)$. *The qualifier "uniform" is appropriate since* $N_i^4(u) = N_0^4(u-i)$; *hence, every blending function is a translate of* N_0^4. *The* r^{th} *segment of a uniform cubic B-spline curve is given by*

$$\mathbf{q}_r(u) = \sum_{i=0}^{3} \mathbf{V}_{r+i} n_i(u), \qquad u \in [0,1] \tag{3.13}$$

where the n*'s are the segments of* N_0^4 *as shown in Figure 3.2.* •

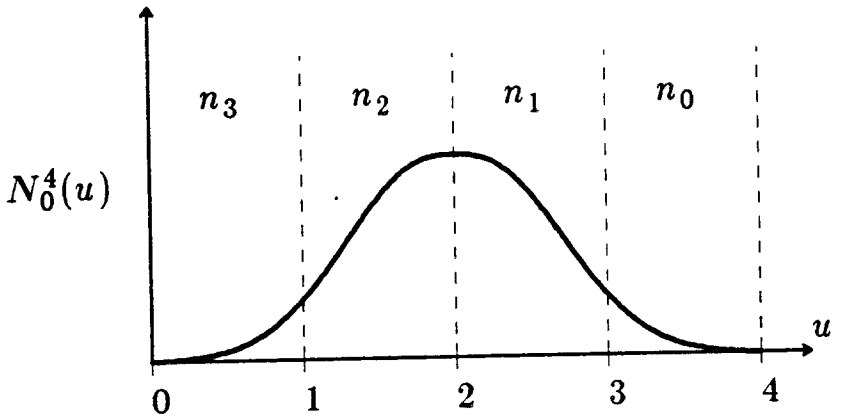

Figure 3.2. *The labeling of the basis segments of the uniform cubic B-spline basis function* $N_0^4(u)$. *The basis segment* $n_i(u)$ *is segment of* $N_0^4(u-i)$ *that is supported on* $[0,1]$. *This labeling scheme causes the indices to increase from right to left, as shown above.*

3.2. Placement of Bézier Vertices

Although Bézier curves are not piecewise, we can construct a spline by stringing together separate Bézier curves, requiring that the curves meet with geometric continuity at each joint. This is the general approach taken by Faux & Pratt [32], Farin [27], Fournier & Barsky [33], and Ramshaw [50].

As a specific example of this process, suppose we are given a cubic Bézier curve $\mathbf{l}$ defined by the control vertices $\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ and are asked to find constraints on the vertices $\mathbf{W}_0, \mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3$ defining a cubic Bézier curve $\mathbf{r}$ such that $\mathbf{l}$ and $\mathbf{r}$ meet with G^2 continuity at the point corresponding to $\mathbf{l}(1)$ and $\mathbf{r}(0)$ (see Figure 3.3).

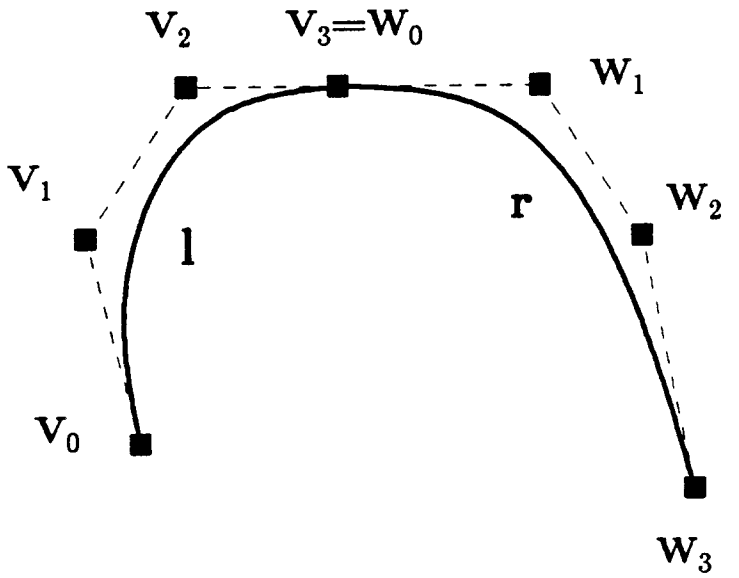

Figure 3.3. *The vertices* $\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ *define a cubic Bézier curve* $\mathbf{l}$*, and the vertices* $\mathbf{W}_0, \mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3$ *define a cubic Bézier curve* $\mathbf{r}$*. We wish to determine constraints on the* $\mathbf{W}$*'s so that* $\mathbf{l}$ *and* $\mathbf{r}$ *meet with* G^2 *continuity at the point* $\mathbf{l}(1) = \mathbf{r}(0)$.

By property 1 of Bézier curves (interpolation of end vertices), $\mathbf{l}(1) = \mathbf{r}(0)$ implies $\mathbf{V}_3 = \mathbf{W}_0$. This establishes G^0 continuity. To establish G^1 continuity, we require that $\mathbf{r}^{(1)}(0) = \beta_1 \mathbf{l}^{(1)}(1)$, which by equation (3.10) implies that

$$3(\mathbf{W}_1 - \mathbf{W}_0) = 3\beta_1(\mathbf{V}_3 - \mathbf{V}_2). \tag{3.14}$$

Using the fact that $\mathbf{W}_0 = \mathbf{V}_3$, (3.14) can be solved for $\mathbf{W}_1$:

$$\mathbf{W}_1 = \mathbf{V}_3 + \beta_1(\mathbf{V}_3 - \mathbf{V}_2). \tag{3.15}$$

Geometrically, equation (3.15) requires $\mathbf{V}_2, \mathbf{V}_3$, and $\mathbf{W}_1$ to be collinear with $\mathbf{V}_3$ between $\mathbf{V}_2$ and $\mathbf{W}_1$. Finally, the G^2 constraint is used to constrain $\mathbf{W}_2$. The resulting expression is

$$\mathbf{W}_2 = \mathbf{V}_3 + (2\beta_1 + \beta_1^2 + \beta_2/2)(\mathbf{V}_3 - \mathbf{V}_2) - \beta_1^2(\mathbf{V}_2 - \mathbf{V}_1) \tag{3.16}$$

Equation (3.16) demands the coplanarity of the vertices $\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3 = \mathbf{W}_0, \mathbf{W}_1$, and $\mathbf{W}_2$. Therefore, the only completely free vertex is $\mathbf{W}_3$. If third order continuity was desired, $\mathbf{W}_3$ would be constrained too. In general, requiring G^r continuity constrains $r + 1$ vertices. The reader is referred to Bartels *et al* [8] or Fournier & Barsky [33] for a more complete treatment.

Remark 3.1: *Strictly speaking, geometric continuity requires that the parametrizations be regular. It is not sufficient to require only that the Beta constraints be satisfied. We mention this here because it is possible for regularity to be violated when using blended polynomial splines.*

For instance, consider the case for cubic Bézier curves. It is always possible to find control vertices $\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ *such that the resulting Bézier curve has an irregularity at any parameter value on* $[0, 1]$. *The reasoning is as follows. Let* u_* *be any real number on* $[0, 1]$, *and consider the parametrization*

$$\mathbf{c}(u) = ((u - u_*)^2, (u - u_*)^3), \qquad u \in [0, 1]. \tag{3.17}$$

Since $\mathbf{c}(u)$ *is a parametric cubic, and since the cubic Bézier basis functions span the cubic polynomials, there must exist vertices* $\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3$ *that describe* $\mathbf{c}(u)$:

$$\mathbf{c}(u) = \sum_{i=0}^{3} \mathbf{V}_i b_i^3(u), \qquad u \in [0, 1]. \tag{3.18}$$

It is easily verified that $\mathbf{c}(u)$ *has an irregularity at* $u = u_*$. *In fact, one can show that* $\mathbf{c}(u)$ *has a discontinuous unit tangent vector at* u_*.

Thus, it is possible to find control vertices that cause a discontinuous unit tangent vector, and therefore violate G^1 *continuity, anywhere on a cubic curve segment. This example points out that if strict geometric continuity is desired, then it is not sufficient to restrict attention to the joints; every point on the interior of the curve segments must also be checked for regularity. In fact, even* $\beta_1 > 0$ *is no guarantee since both curves may be irregular at the joint.*

3.3. Beta-spline Curves

In his thesis, Barsky [3] introduced the cubic Beta-spline curve technique. The cubic Beta-spline is the geometrically continuous analog of the uniform cubic B-spline from Example 3.1. The Beta-spline basis segments $b_i(\beta_1, \beta_2; u)$, $i = 0, 1, 2, 3$, are used in place of the B-spline basis segments $n_i(u)$ in equation (3.13), and are derived by requiring the curve segments of the spline to meet with G^2 rather than C^2 continuity. Specifically,

$$\begin{aligned} \mathbf{q}_r(0) &= \mathbf{q}_{r-1}(1) \\ \mathbf{q}_r^{(1)}(0) &= \beta_1 \mathbf{q}_{r-1}^{(1)}(1) \\ \mathbf{q}_r^{(2)}(0) &= \beta_1^2 \mathbf{q}_{r-1}^{(2)}(1) + \beta_2 \mathbf{q}_{r-1}^{(1)}(1). \end{aligned} \tag{3.19}$$

If these constraints are to hold for any choice of the control vertices, it must be that [3,4,5]

$$\begin{aligned} 0 &= b_0(\beta_1, \beta_2; 1) \\ b_0(\beta_1, \beta_2; 0) &= b_1(\beta_1, \beta_2; 1) \\ b_1(\beta_1, \beta_2; 0) &= b_2(\beta_1, \beta_2; 1) \\ b_2(\beta_1, \beta_2; 0) &= b_3(\beta_1, \beta_2; 1) \\ b_3(\beta_1, \beta_2; 0) &= 0 \end{aligned} \tag{3.20}$$

$$\begin{aligned} 0 &= \beta_1 b_0^{(1)}(\beta_1, \beta_2; 1) \\ b_0^{(1)}(\beta_1, \beta_2; 0) &= \beta_1 b_1^{(1)}(\beta_1, \beta_2; 1) \\ b_1^{(1)}(\beta_1, \beta_2; 0) &= \beta_1 b_2^{(1)}(\beta_1, \beta_2; 1) \\ b_2^{(1)}(\beta_1, \beta_2; 0) &= \beta_1 b_3^{(1)}(\beta_1, \beta_2; 1) \\ b_3^{(1)}(\beta_1, \beta_2; 0) &= 0 \end{aligned} \tag{3.21}$$

$$\begin{aligned} 0 &= \beta_1^2 b_0^{(2)}(\beta_1, \beta_2; 1) + \beta_2 b_0^{(2)}(\beta_1, \beta_2; 1) \\ b_0^{(2)}(\beta_1, \beta_2; 0) &= \beta_1^2 b_0^{(2)}(\beta_1, \beta_2; 1) + \beta_2 b_0^{(2)}(\beta_1, \beta_2; 1) \\ b_1^{(2)}(\beta_1, \beta_2; 0) &= \beta_1^2 b_2^{(2)}(\beta_1, \beta_2; 1) + \beta_2 b_2^{(2)}(\beta_1, \beta_2; 1) \\ b_2^{(2)}(\beta_1, \beta_2; 0) &= \beta_1^2 b_3^{(2)}(\beta_1, \beta_2; 1) + \beta_2 b_3^{(2)}(\beta_1, \beta_2; 1) \\ b_3^{(2)}(\beta_1, \beta_2; 0) &= 0 \end{aligned} \tag{3.22}$$

$$b_0(\beta_1, \beta_2; 0) + b_1(\beta_1, \beta_2; 0) + b_2(\beta_1, \beta_2; 0) b_3(\beta_1, \beta_2; 0) = 1. \tag{3.23}$$

where equation (3.23) has been chosen so that the basis segments form a partition of unity in the sense of equation (3.6). The above system of equations can

be solved symbolically using a symbolic algebra system such as Vaxima [31] or REDUCE [40,42]. The resulting basis segments can be found in [3,4,5].

The Beta-spline described above is said to be *uniformly-shaped.* That is, the shape parameters $\beta_{1,j}$ and $\beta_{2,j}$ at the j^{th} joint are required to be equal to a global set β_1 and β_2:

$$\begin{aligned} \beta_{1,j} &= \beta_1 \\ \beta_{2,j} &= \beta_2 \end{aligned} \qquad \forall j. \tag{3.24}$$

Work by Barsky [3], and later work by Barsky & Beatty [5], developed a method of assigning shape parameter values locally, but their technique resulted in a rational polynomial representation of high degree. Bartels & Beatty [9], and independently Goodman [37], have discovered a cubic representation with local shape parameters. These representations are said to be *locally-shaped.*

In keeping with the notion of Beta-splines as geometrically continuous B-splines, we can define the space of geometrically continuous piecewise polynomials. The Beta-spline basis functions can then be viewed as the spanning set of least support. Goodman [37] has verified the existence of such a spanning set, and has found that the support is identical to the support of the corresponding B-spline basis. Unfortunately, his proof is not a constructive one, so the process of determining the Beta-spline basis is still an *ad hoc* one. In this more general context, the Beta-spline introduced by Barsky is most conveniently termed the *uniformly-shaped cubic Beta-spline.*

It would indeed be interesting (and useful) to develop an evaluation algorithm for locally shaped Beta-splines of arbitrary degree. It would also be useful to construct an algorithm for inserting new knots into Beta-spline curves. In its most general form, such an algorithm would be the geometrically continuous analog of the *Oslo Algorithm* [19] for knot insertion into B-splines of arbitrary order.

3.4. Geometrically Continuous Catmull-Rom Splines

The Beta-spline presented in the previous section is an approximating technique possessing shape parameters. Although the Beta-spline technique has proven to be useful in *a priori* design, there are many applications where an interpolating spline is required. It is useful to investigate the use of geometric continuity, and hence shape parameters, for interpolating splines.

The Wilson-Fowler spline [34] is an interpolating polynomial technique requiring continuity of unit tangent vectors and approximate continuity of curvature

vectors. The ν-spline, due to Nielson [49], is an interpolating polynomial technique requiring continuity of first parametric derivative and curvature vectors (this corresponds to $\beta_1 = 1$, with β_2 being arbitrary). However, the Wilson-Fowler spline and the ν-spline do not have the property of local control. On the other hand, the Catmull-Rom splines [18] have local control and can be either interpolating or approximating. However, the splines Catmull & Rom explicitly construct are parametrically continuous.

DeRose & Barsky [23] combined the notion of geometric continuity with the work of Catmull & Rom to describe the class of *Geometrically Continuous Catmull-Rom Splines*. Members of this class are either interpolating or approximating, have local control, and by virtue of geometric continuity, possess shape parameters. A member of the class is constructed by combining Beta-spline blending functions with a geometrically continuous extension of the classical *Lagrange polynomials*.

The interpolating members of the class may be useful in design situations where interpolating splines with local control are desirable. The additional flexibility provided by the shape parameters allow local modifications to "tweak" the design curve. It may also be possible to use the interpolating members to construct *transfinite methods* such as geometrically continuous *Gordon surfaces* [38].

3.5. Summary

Examples of the use of geometric continuity and the univariate Beta constraints were presented. These included the placement of Bézier vertices, the construction of the uniform cubic Beta-spline, a discussion of locally shaped Beta-splines of arbitrary order, and a brief discussion of the class of geometrically continuous Catmull-Rom splines.

It should be emphasized that the construction of geometrically continuous techniques is still an *ad hoc* process. For example, if it is known that a G^3 approximating spline is needed for a particular application, it is possible to set up and symbolically solve a linear system of equations to describe the basis segments. However, the G^3 solution does not aid in the construction of a G^4 technique. To derive a G^4 technique, one must currently revert to setting up and solving a new (and larger) system of equations. It is hoped that a general algorithm for constructing Beta-splines of arbitrary order will soon be developed. Such an algorithm may also provide the key to an algorithm for inserting new knots into Beta-splines of arbitrary order.

The characterizations of geometric continuity given in Chapters 2 and 6 describe the number of shape parameters and their allowable values, but say nothing about how the curve will deform in response to a change in a shape parameter. The effect of changing a shape parameter must be determined empirically — no theorems currently exist to describe how changing a given shape parameter will perturb the curve. Indeed, preliminary results suggest that the effect of a particular shape parameter may depend on the polynomial order of the curve. It is also possible that the effect is dependent on the interpolating or approximating nature of the technique.

4

Tensor Product Surfaces

In this chapter, we examine the construction of tensor product surfaces from geometrically continuous univariate basis functions. It is shown that a tensor product of univariate geometrically continuous basis functions results in a special case of a geometrically continuous surface.

4.1. Introduction

A spline surface can be constructed from univariate basis functions as follows. Let $\mathbf{V}_{i,j}$, $i = 0, ..., m$ and $j = 0, ..., n$ be a rectangular grid of control vertices, and let $W_i(u)$, $u \in [u_0, u_f]$, $W_j(v)$, $v \in [v_0, v_f]$ be basis functions as in equation (3.1). A *tensor product surface* is generated by a bivariate parametrization $\mathbf{S}(u, v)$ of the form

$$\mathbf{S}(u,v) = \sum_{i=0}^{m} \sum_{j=0}^{n} \mathbf{V}_{i,j} W_i(u) W_j(v) \tag{4.1}$$

$$(u,v) \in [u_0, u_f] \times [v_0, v_f].$$

In a sense, tensor product surfaces are constructed using "curve technology". That is, the bivariate blending function is constructed by forming a product of two univariate functions, as shown in equation (4.1). One reason that tensor product surfaces are so common is that parametrically smooth univariate blending functions result in a parametrically smooth tensor product surface. Formally:

Lemma 4.1: *If $W_i(u)$ and $W_j(v)$ are C^r, then $\mathbf{S}(u, v)$ is C^r.*

Proof: Follows immediately from the definition of C^r of a bivariate function and the tensor product form of $\mathbf{S}$. ∎

Example 4.1: *Tensor Product B-spline: The tensor product B-spline surface is of the form of equation (4.1), where the blending functions are the B-spline basis functions from equation (3.11) (cf. Gordon & Riesenfeld [39]).* •

4.2. Geometric Continuity of Tensor Product Surfaces

The next theorem goes one step further than Lemma 4.1; it shows that a tensor product of any geometrically continuous univariate blending functions results in a geometrically continuous surface.

Theorem 4.1: *Let* $\mathbf{S}$ *be as in equation (4.1) and regular. If* $W_i(u)$ *and* $W_j(v)$ *are each (independently)* G^r*, then* $\mathbf{S}$ *is* G^r*.*

Proof: Since the basis segments of $W_i(u)$ satisfy the Beta constraints, they can be reparametrized to obtain basis segments that meet with derivative continuity. Thus, it is possible to reparametrize $W_i(u)$ to obtain a C^r basis function $\widetilde{W}_i(\widetilde{u})$. Of course, the same is true of $W_j(v)$. Lemma 4.1 can then be invoked using $\widetilde{W}_i(\widetilde{u})$ and $\widetilde{W}_j(\widetilde{v})$ to show that $\widetilde{\mathbf{S}}$ is C^r, where $\widetilde{\mathbf{S}}$ is defined as

$$\widetilde{\mathbf{S}}(\widetilde{u},\widetilde{v}) = \sum_{i=0}^{m}\sum_{j=0}^{n} \mathbf{V}_{i,j}\, \widetilde{W}_i(\widetilde{u})\widetilde{W}_j(\widetilde{v}) \qquad (4.2)$$

$$(\widetilde{u},\widetilde{v}) \in [\widetilde{u}_0,\widetilde{u}_f] \times [\widetilde{v}_0,\widetilde{v}_f].$$

∎

Example 4.2: *The tensor product Beta-spline is an example of a geometrically continuous tensor product technique. The effect of shape parameter change is shown in Figure 4.1.*

Tensor product spline surfaces can also be constructed by constraining the vertices of separate tensor product Bézier surface patches. For details concerning G^1 *and* G^2 *continuity, see Veron et al [60], Faux & Pratt [32], and Kahmann [44].*

•

Since G^r univariate basis functions are used in each of u and v directions, $2r$ shape parameters are possible in a G^r tensor product surface. However, the shape parameters are constant along the patch boundaries.

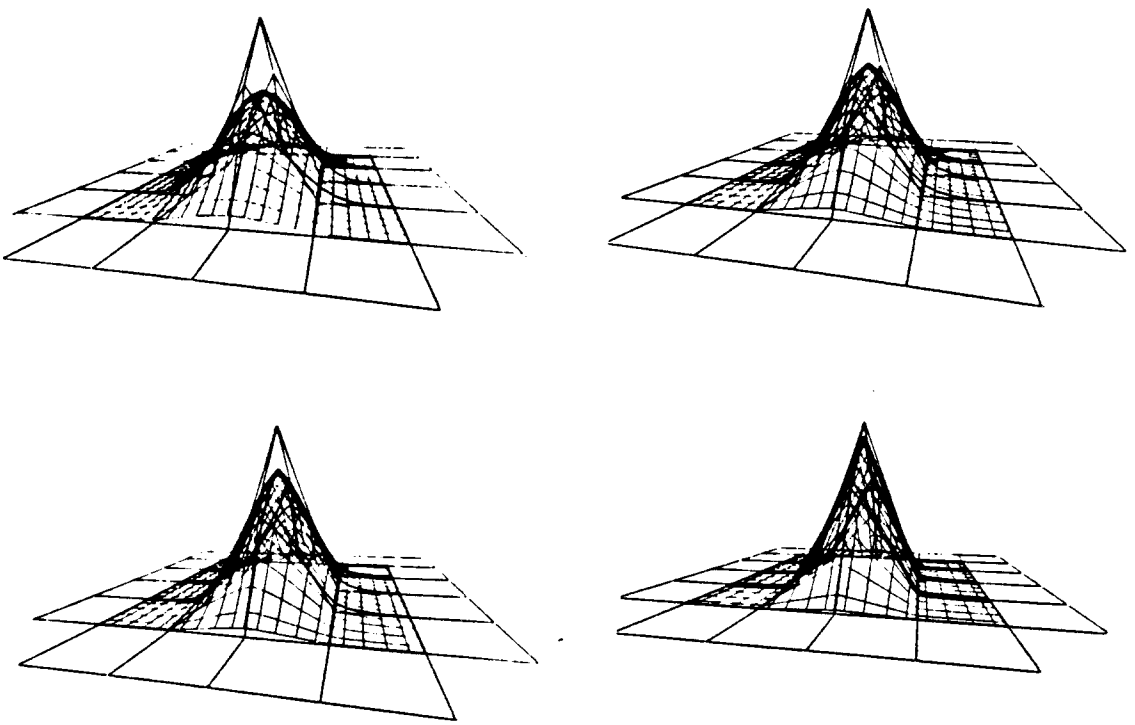

Figure 4.1. *This set of bicubic Beta-spline surfaces shows the effect of increasing the value of the shape parameter β_2. The values of β_2 are 0 for the top left surface, 5 for the top right surface, 10 for the bottom left surface, and 50 for the surface in the bottom right position.*

4.3. Summary

Geometric continuity for tensor product surfaces is not very interesting from a theoretic standpoint. It follows directly from geometric continuity of the univariate blending functions used to construct the tensor product. However, geometrically continuous tensor product forms are interesting from a practical standpoint, as demonstrated by the recent enthusiasm over the tensor product bicubic Beta-spline of Example 4.2.

5

Triangular Spline Surfaces

In this chapter, examples of the use of the bivariate Beta constraints for spline surfaces comprised of triangular patches are presented. These include the placement of control vertices for Bézier triangles, and the construction of the G^1 analog of the triangular cubic B-spline, called the triangular cubic Beta-spline.

5.1. Introduction

One of the primary reasons that tensor product surfaces are prevalent in CAGD systems is the ease with which they can be implemented and theoretically characterized. For instance, it was shown in Chapter 4 that smoothness of a tensor product surface follows directly from smoothness of the univariate basis functions from which the bivariate basis functions are constructed. Despite their simplicity, tensor product forms can be quite restrictive in the sense that the bivariate basis function is required to factor into a product of two univariate functions. Of course, there are many smooth bivariate functions that do not factor.

Recent work by researchers such as Barnhill, Böhm, Farin, Kahmann, Micchelli, and others, has begun to popularize a class of non-tensor product surfaces known as *triangular splines*. These splines deserve the name "triangular" because the patches are defined on triangular, rather than rectangular, regions of the parameter plane. Triangular surfaces offer several advantages over tensor products, including:

- *Lower degree:* Triangular forms often have lower degree when compared to

the tensor product form having equal order of continuity. For instance, a C^2 tensor product B-spline is a bivariate polynomial of degree 9, while the C^2 triangular B-spline [53], is a bivariate polynomial of degree 4. Moreover, the disparity in degree grows rapidly with increasing order of continuity [53].

- *More locality:* Triangular forms are often more local than the corresponding tensor product form. For instance, modification of a control vertex in the C^2 triangular B-spline affects 24 triangular patches. As will be shown in Section 5.5, each triangle can be viewed as half a rectangle, implying that 12 rectangular patches are affected. This is as compared to the 16 rectangular patches that would be altered by modification of a control vertex in a C^2 tensor product B-spline.
- *Scattered Data Interpolation:* The problem of fitting smooth surfaces to data scattered in two and three dimensions is often most easily solved using patches of triangular topology [2,30]. Farin [30] has shown that the triangular Bézier surface can be an effective tool in this type of application. While it is possible to use degenerate (irregular) rectangular patches, the degeneracies may wreak havoc on algorithms manipulating the representation. Indeed, this was one of the reasons that continuity of irregular patches was avoided in Chapter 2.

For the reasons cited above, it is believed that triangular splines will play an increasingly important role in CAGD in the near future. It is therefore of interest to study geometric continuity for this class of surfaces.

5.2. Notation

Before beginning the discussion of triangular surfaces, we define some convenient notation. Diacritical vectors will be used to denote tuples with two or three coordinates; in particular $\vec{i} = (i_1, i_2, i_3)$, $\vec{u} = (u_1, u_2, u_3)$, and $\vec{g} = (g_1, g_2)$. As in Section 1.2, the norm of such a tuple is defined to be the sum of the components.

Let T denote a triangle with vertices $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$. An arbitrary point $\mathbf{t}$ in T has a unique representation $\vec{u} = (u_1, u_2, u_3)$ according to

$$\mathbf{t} = u_1\mathbf{t}_1 + u_2\mathbf{t}_2 + u_3\mathbf{t}_3 \tag{5.1}$$

where $u_1, u_2, u_3 \in \Re_+$ and $|\vec{u}| = u_1 + u_2 + u_3 = 1$. $\mathbf{t}$ is said to have *barycentric coordinates* $\vec{u}$ relative to T.

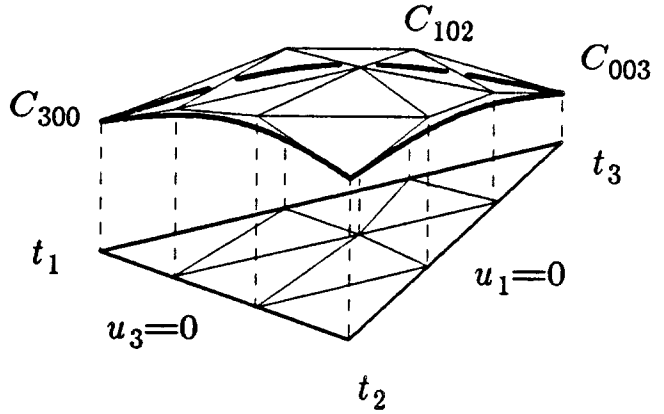

Figure 5.1. *The geometric relationship between the graph of p and the corresponding Bézier net.*

5.3. Triangular Bézier Surfaces

In Section 3.1.1, the Bézier basis functions were introduced. The definition of these basis functions from equation (3.8) can be rewritten as

$$b^d_{i_1}(u) = b^d_{i_1,i_2}(u_1,u_2) = \frac{d!}{i_1!\,i_2!}u_1^{i_1}\,u_2^{i_2} \tag{5.2}$$

where $u_1, u_2 \in \Re$, $u_1 + u_2 = 1$, $i_1, i_2 \in \mathbf{Z}_+$, and $i_1 + i_2 = d$. de Casteljau [17], and later Sabin [53], extended this form to the bivariate case to define the *bivariate Bézier basis functions*:

$$b^d_{\vec{i}}(\vec{u}) = \frac{d!}{i_1!\,i_2!\,i_3!}u_1^{i_1}\,u_2^{i_2}\,u_3^{i_3}, \tag{5.3}$$

where $|\vec{i}| = d$, $|\vec{u}| = 1$, and $u_1, u_2, u_3 \in \Re_+$.

The bivariate Bézier basis functions (also called the bivariate Bernstein polynomials) of degree d can be shown to span the bivariate polynomials of degree d, implying that any polynomial p of degree less than or equal to d can be uniquely expressed as

$$p(\vec{u}) = \sum_{|\vec{i}|=d} C_{\vec{i}}\,b^d_{\vec{i}}(\vec{u}), \qquad C_{\vec{i}} \in \Re. \tag{5.4}$$

The *Bézier coefficients* $C_{\vec{i}}$ that describe p have a geometric interpretation relative to the graph of p. If (u_1, u_2, u_3) are barycentric coordinates relative to a triangle T with vertices $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$, then the graph of p over T is described by the parametrization $\mathbf{p}(\vec{u}) = (\vec{u}, p(\vec{u}))$, $u_1, u_2, u_3 \in \Re_+$; the points $(\frac{1}{d}\vec{i}, C_{\vec{i}})$ form a triangulated control net (called a *Bézier net*) that mimics the shape of $\mathbf{p}(\vec{u})$ (see Figure 5.1).

The Bézier representation of a bivariate polynomial has some interesting and useful properties, including:

1) *Convex hull property:* The graph of p is guaranteed to lie in the convex hull of the Bézier net.
2) *Boundary curves:* Consider the boundary curve corresponding to $\mathbf{p}$ restricted to $u_1 = 0$. This curve is completely determined by the coefficients $C_{(0,i_1,i_2)}$. In general, boundary curves are determined by the Bézier coefficients along that boundary.
3) *Directional derivatives:* Let $\vec{A}^1$ and $\vec{A}^2$ be the barycentric coordinates of two points, and let $\vec{A} = \vec{A}^2 - \vec{A}^1$ be their vector difference. Farin [30] shows that the directional derivative of p in the direction of $\vec{A}$, denoted $D_{\vec{A}}p$, is given by

$$D_{\vec{A}}p(\vec{u}) = d \sum_{|\vec{i}|=d-1} C_{\vec{i}}^{[1]} b_{\vec{i}}^{d-1}(\vec{u}) \tag{5.5}$$

where

$$C_{\vec{i}}^{[1]} = \sum_{r=1}^{3} A_r C_{\vec{i}+\vec{j}_r} \tag{5.6}$$

and A_r denotes the r^{th} component of $\vec{A}$.

Example 5.1: *As a specific example of the use of equations (5.5) and (5.6), let p be a cubic polynomial, and let $\vec{A} = (2/3, 1/3, 0) - (1, 0, 0) = (-1/3, 1/3, 0)$, a horizontal vector pointing to the right. The value of $D_{\vec{A}}p(\vec{u})$ is the slope of p in the direction $\vec{A}$, evaluated at the point $\vec{u}$. The Bézier net for the graph of $D_{\vec{A}}p$ can be computed from equation (5.5). The computation is shown graphically in Figure 5.2.* •

The Bézier basis functions can be used to construct parametric surface patches called *Bézier triangles.* A Bézier triangle $\mathbf{p}$ of degree d on a triangular domain T, is constructed by replacing the Bézier coefficients $C_{\vec{i}}$ in equation (5.4) with control vertices $\mathbf{C}_{\vec{i}}$ chosen from $\Re^3$:

$$\mathbf{p}(\vec{u}) = \sum_{|\vec{i}|=d} \mathbf{C}_{\vec{i}} b_{\vec{i}}^{d}(\vec{u}), \qquad u_1, u_2, u_3 \in \Re_+, \quad |\vec{u}| = 1. \tag{5.7}$$

Not surprisingly, the collection of control vertices is called the Bézier net for $\mathbf{p}$. Bézier triangles possess the three properties of Bézier representations (ennumerated above) where the Bézier coefficients $C_{\vec{i}}$ are replaced by the control

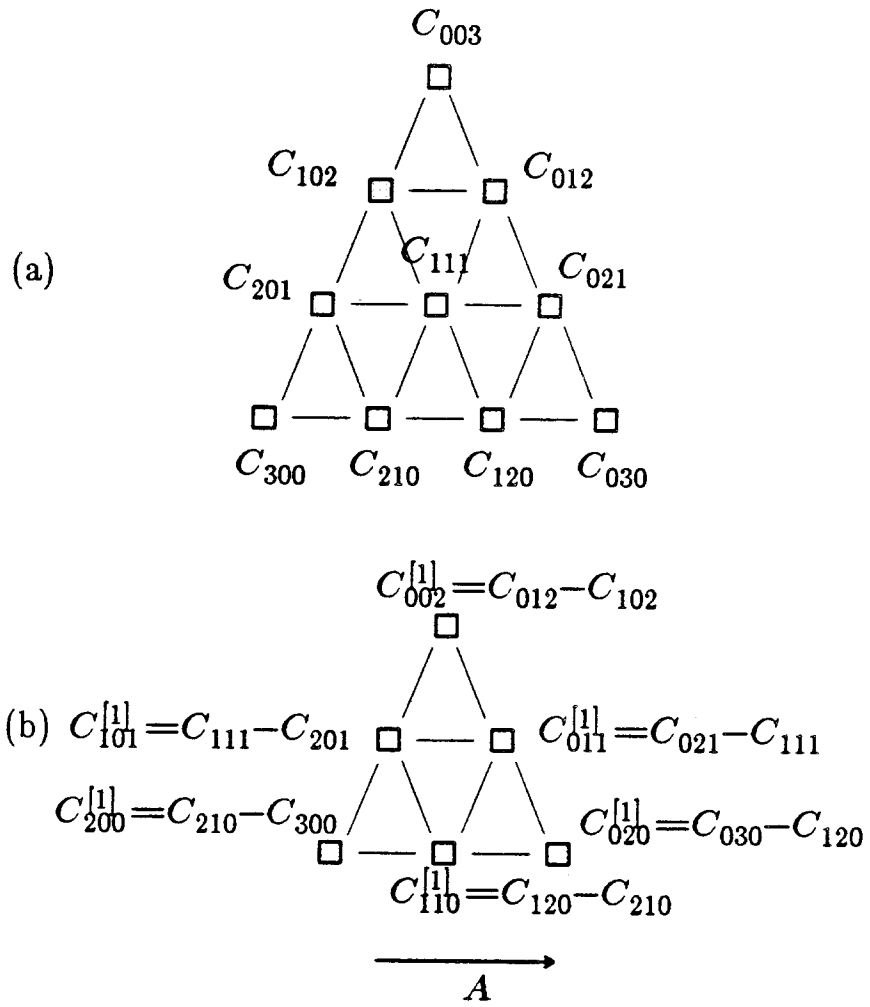

Figure 5.2. *Figure (a) is a view from above the parameter plane of the Bézier net for $p(\vec{u})$ as defined in equation (5.4), with $d = 3$. Figure (b) shows the Bézier net for the directional derivative surface $D_{\vec{A}}p(\vec{u})$, where $\vec{A}$ is a horizontal vector pointing to the right.*

vertices $\mathbf{C}_{\vec{i}}$. In fact, we've already seen an example of a Bézier triangle: the graph of the bivariate function $p(\vec{u})$ from equation (5.4) is the Bézier triangle defined by the net $(\frac{1}{d}\vec{i}, C_{\vec{i}})$.

Just as Bézier curves can be pieced together to create a spline curve, Bézier triangles can be pieced together to create a spline surface. For instance, suppose that $\mathbf{A}(u_1, u_2, u_3)$ and $\mathbf{B}(w_1, w_2, w_3)$ are two cubic Bézier triangles to be stitched together with G^1 continuity along the $u_1 = 0$ and $w_1 = 0$ boundaries. If $\mathbf{a}_{\vec{i}}$ and $\mathbf{b}_{\vec{i}}$ denote the Bézier nets for $\mathbf{A}$ and $\mathbf{B}$, respectively, then C^0 continuity requires

that

$$\mathbf{b}_{(0,j,3-j)} = \mathbf{a}_{(0,j,3-j)}, \qquad j = 0,1,2,3. \tag{5.8}$$

If G^1 continuity is desired, then $\mathbf{A}$ and $\mathbf{B}$ must also satisfy [44]

$$\mathbf{b}_{(1,j,2-j)} = \lambda\, \mathbf{a}_{(0,j,3-j)} + \eta\, \mathbf{a}_{(0,j+1,2-j)} + \nu\, \mathbf{a}_{(1,j,2-j)} \qquad j = 0,1,2 \tag{5.9}$$

where $\lambda, \eta, \nu \in \Re$ are arbitrary, but subject to $\lambda+\eta+\nu = 1$. The reader is referred to Farin [28,29,30] or Kahmann [44] for a more complete treatment of first and second order geometric continuity between Bézier triangles.

5.4. Triangular B-splines

In his thesis, Sabin [53] introduced the triangular B-spline technique as well as the triangular Bézier surfaces of Section 5.3. Roughly speaking, a triangular B-spline surface of polynomial order k, $\mathbf{S}(u,v)$, is defined by a control net $< \mathbf{V}_{i_1,i_2} >$ of regular triangular topology (see Figure 5.3) according to

$$\mathbf{S}(u,v) = \sum_{i_1,i_2} \mathbf{V}_{i_1,i_2}\, B^k_{i_1,i_2}(u,v). \tag{5.10}$$

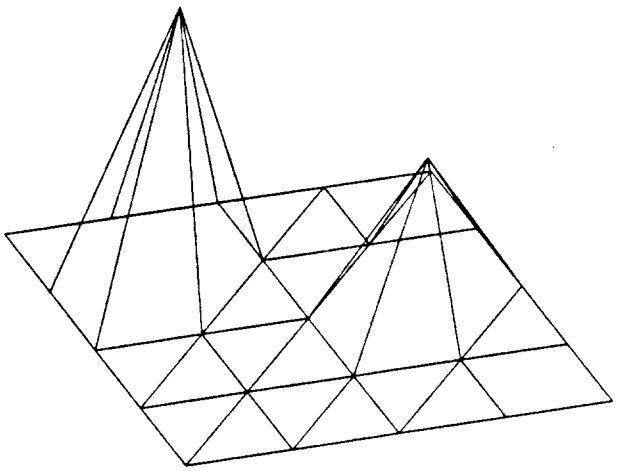

Figure 5.3. *A control net with regular triangular topology.*

The blending functions $B^k_{i_1,i_2}(u,v)$ are piecewise polynomial functions of order k defined on a regular triangulation of the parameter plane. Because of

the regular nature of the triangulation, the blending functions are translates of a canonical function $B^k(u,v)$, called the *triangular B-spline basis function of order k*. This situation is analogous to the uniform B-splines of Section 3.1.2. Sabin [53] originally defined the basis functions as a convolution of lower degree B-splines. It is now common to define these splines as shadows of higher dimensional polytopes [12,22,43].

Example 5.2: *Triangular Cubic B-spline: As a specific example of a triangular B-spline, consider the case of bivariate cubics (i.e., degree 3 bivariate polynomials). The cubic B-spline basis function has the characteristic "hump" shape of an approximating technique as shown in Figure 5.4. Each triangular patch of the basis function, called a basis patch, is a bivariate cubic polynomial (cf. Sabin [53] or Böhm [10]). The patches are constructed so that a spline surface* $\mathbf{S}$ *as in equation (5.10) is guaranteed to be* C^1 *continuous (see Figure 5.5). The basis function is non-negative and normalized so that the surface is independent of the coordinate system in which the control vertices are expressed. The surface is therefore guaranteed to lie in the convex hull of the control net. Referring to Figure 5.4, the basis function is locally supported on 13 triangular subdomains of the parameter plane, implying that perturbation of a control vertex modifies only 13 patches of the surface.* •

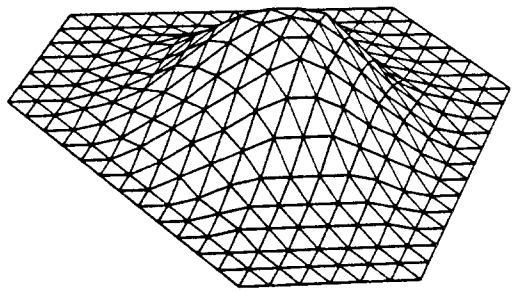

Figure 5.4. *The triangular cubic B-spline basis function is plotted above.*

5.5. Triangular Beta-spline Surfaces

Because of the advantages of triangular spline surfaces, it is desirable to extend the triangular B-splines to obtain their geometrically continuous counterparts the *triangular Beta-splines.* One would like to have a general evaluation algorithm for

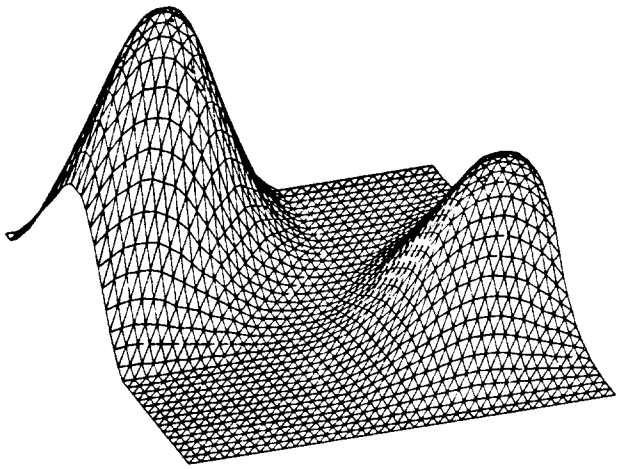

Figure 5.5. *The triangular cubic B-spline surface defined by the control net of Figure 5.3.*

arbitrary degree, but that has not yet been found. Instead, we present an *ad hoc* construction of the *triangular cubic Beta-spline.*

The triangular cubic Beta-spline technique is characterized by the triangular cubic Beta-spline basis function $N(u,v)$. Assuming N has the same support as the corresponding B-spline basis function $B^3(u,v)$, we seek expressions for the 13 basis patches that comprise N, constructed so that the resulting spline surface

$$\mathbf{S}(u,v) = \sum_{i_1,i_2} \mathbf{V}_{i_1,i_2}\, N(u-i_1, v-i_2) \tag{5.11}$$

is guaranteed to be G^1 continuous.

Each basis patch is assumed to be a bivariate cubic polynomial, and is therefore determined by 10 coefficients. Thus, for the 13 patches comprising N, there are a total of 130 unknowns. We could approach the problem in a manner analogous to the univariate cubic Beta-spline basis segments presented in Section 3.3, but that would require the symbolic solution of a system of 130 simultaneous linear equations. This seemingly dismal situation can be drastically improved by expressing the basis patches in terms of the bivariate Bézier basis (see Section 5.3). This process of *Bézier reduction* allows the system to be reduced to 25 unknowns. Bézier reduction could have been used to construct the basis segments for cubic Beta-spline curves, but the reduction would have been much less dramatic: 16 unknowns for the naive method, versus 9 for Bézier reduction.

5.5.1. Derivation of the Triangular Cubic Beta-spline Basis Patches

The triangular cubic B-spline is a C^1 technique with local control, so the triangular cubic Beta-spline should be a G^1 representation with local control. The basis patches must be constructed so that:

1) N vanishes around its perimeter.
2) The tangent plane of N around the perimeter must coincide with the (u, v) plane.
3) Interior boundaries satisfy the bivariate Beta constraints.
4) The basis patches must form a partition of unity.

Constraint 1 is necessary for local support, or equivalently, for local control. Constraints 2 and 3 combine to ensure the G^1 character of the spline surface $\mathbf{S}(u, v)$. Finally, Constraint 4 is necessary (and sufficient) for $\mathbf{S}(u, v)$ to be coordinate system independent.

The first step in the construction of the basis function N is the partitioning of the parameter plane into a set of regular triangles, as shown in Figure 5.6. A global coordinate system on the plane is established such that edges of the triangular domains are one unit in length (see Figure 5.6).

As suggested by equation (5.11), with each *lattice point* (i_1, i_2) we associate a blending function. To do this, we choose a *reference point* for the canonical blending function $N(u, v)$; the blending function $N(u - i_1, v - i_2)$, associated with the point (i_1, i_2) in the parameter plane, is then defined to be a copy of the canonical blending function with its reference point at (i_1, i_2) (see Figure 5.7).

Referring to Figure 5.7, the basis patches of N naturally fall into two classes: those pointing upward, called "up" patches, and those pointing downward, called "down" patches. Let superscript $\uparrow$ denote quantities corresponding to up patches and superscript $\downarrow$ denote quantities corresponding to down patches; the symbol $\uparrow\downarrow$ will be used to denote either up or down, much like $\pm$. In addition to the global (u, v) coordinate system, it is convenient to set up a local barycentric coordinate system on each patch as shown in Figure 5.8.

Let $n^{\uparrow\downarrow}_{k_1,k_2}(\vec{u})$ denote a generic basis patch, $\vec{u}$ a barycentric coordinate relative to the domain triangle where of the patch. We choose an indexing scheme as show in Figure 5.9 using the following rationale. The blending function $N(u, v)$ is used to weight the vertex $\mathbf{V}_{0,0}$ in the spline surface, as dictated by equation (5.11). The central patch of $N(u, v)$ is the one that is (in some sense) most strongly

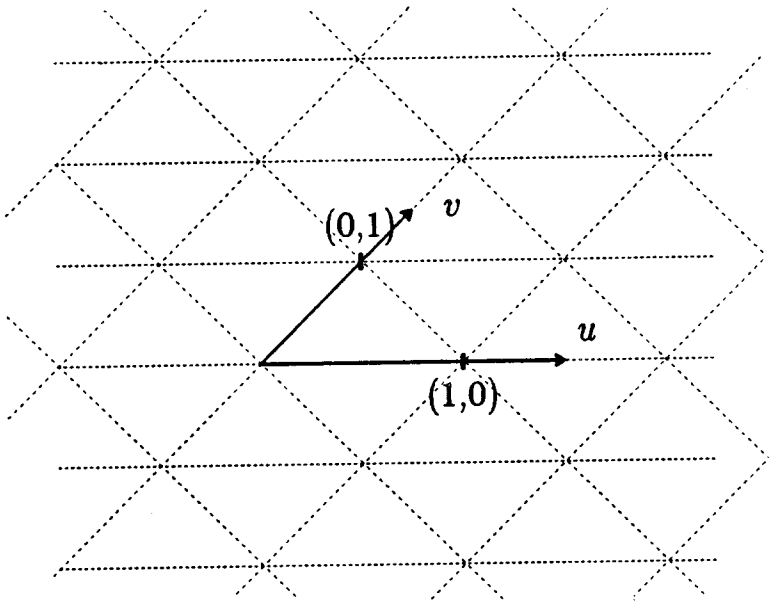

Figure 5.6. *The parameter plane is partitioned into a set of regular triangles one unit in length on each edge. A global coordinate system (u, v) is set up as shown above.*

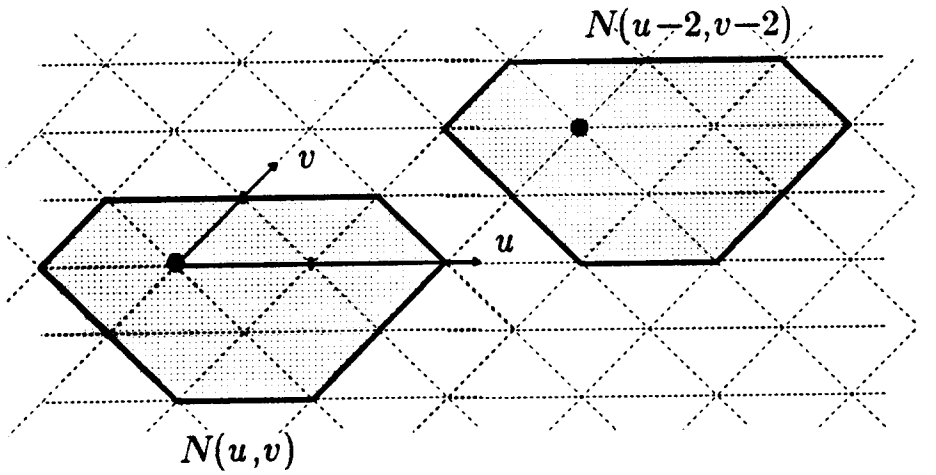

Figure 5.7. *The blending function $N(u, v)$ is placed in the parameter plane with its reference point (shown as a black circle) located at the origin. The blending function associated with the lattice point $(2, 2)$ is $N(u-2, v-2)$, which can be thought of as a copy of $N(u, v)$ with its reference point at $(2, 2)$, as shown above.*

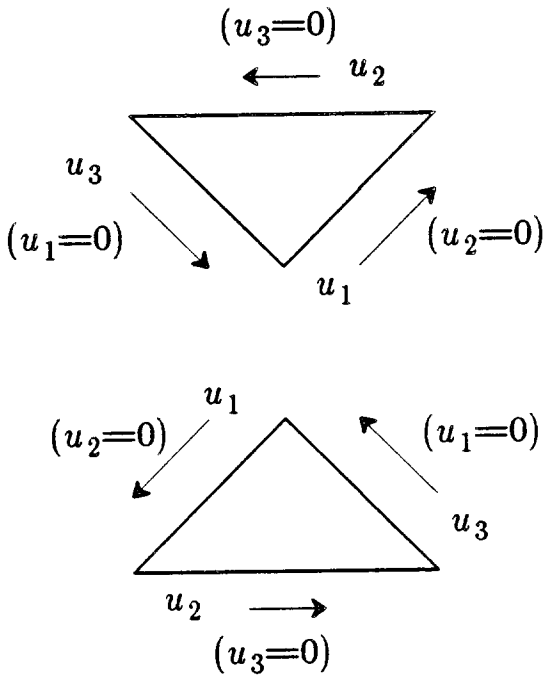

Figure 5.8. *The local barycentric coordinate systems for up and down basis patches.*

related to the vertex $\mathbf{V}_{0,0}$, so we choose to label it with a 0,0 index. Let us fix attention on the domain triangle on which $n^{\downarrow}_{0,0}$ is defined, and consider the blending function $N(u-1,v)$ that is associated with the control vertex $\mathbf{V}_{1,0}$. This triangle is darkened in Figure 5.10. The part of $N(u-1,v)$ that lies above the darkened triangle is given a subscript of $(1,0)$ to reflect the fact that it is associated with the vertex $\mathbf{V}_{1,0}$. Similarly, the patch labeled $n^{\downarrow}_{1,-1}$ is that part of $N(u-1,v+1)$ (the blending function associated with $\mathbf{V}_{1,-1}$) that lies above the darkened triangle. In general, the down patch with index i_1, i_2 is part of $N(u-i_1, v-i_2)$ supported over the darkened triangle. Similarly, the up patch with index i_1, i_2 is the part of $N(u-i_1, v-i_2)$ supported over the up triangle whose apex is at the origin (any other up triangle adjacent to the darkened one would serve equally as well).

As mentioned previously, we express each basis patch $n^{\uparrow\downarrow}_{k_1,k_2}(\vec{u})$ in terms of cubic Bézier coefficients $f^{\uparrow\downarrow}_{k_1,k_2,\vec{i}}$. Explicitly,

$$n^{\uparrow\downarrow}_{k_1,k_2}(\vec{u}) = \sum_{|\vec{i}|=3} f^{\uparrow\downarrow}_{k_1,k_2,\vec{i}} b^3_{\vec{i}}(\vec{u}) \tag{5.12}$$

where $b^3_{\vec{i}}(\vec{u})$ denotes a bivariate cubic Bézier basis function from equation (5.3).

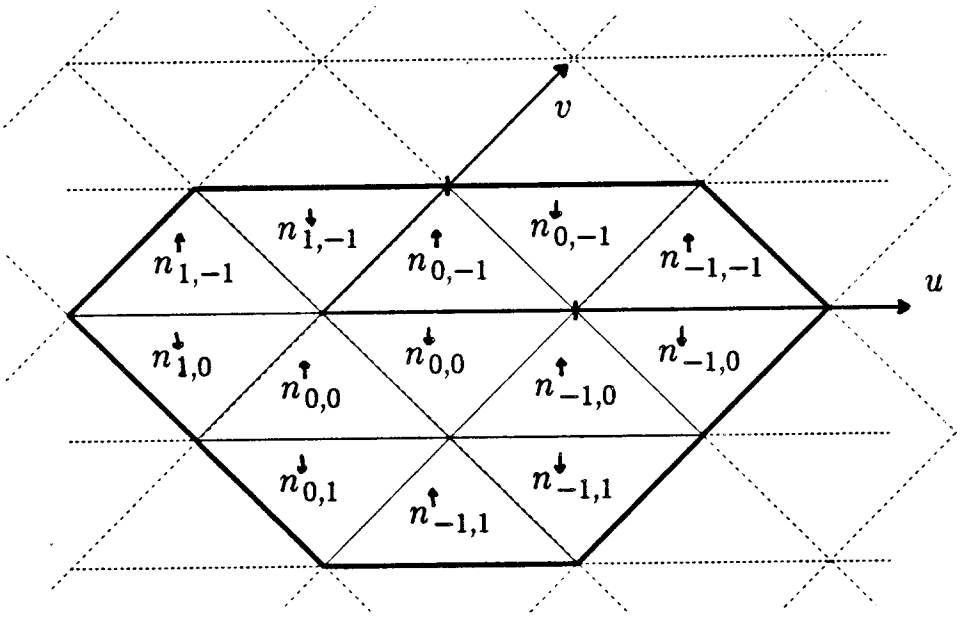

Figure 5.9. *The indexing of the basis patches of $N(u,v)$ is as shown above.*

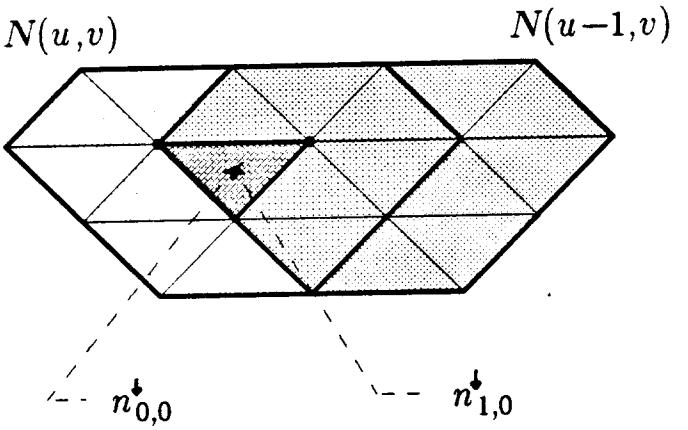

Figure 5.10. *A view from above the (u,v) plane showing the supports of $N(u,v)$ and $N(u-1,v)$. $N(u,v)$ has its support lightly shaded; the support of $N(u-1,v)$ is slightly darker. The basis patch of $N(u,v)$ supported on the dark triangle is given an index of $0,0$. Similarly, the basis patch of $N(u-1,v)$ supported on the dark triangle is given an index of $1,0$.*

Referring to Figure 5.11, we seek the Bézier coefficients $f^{\uparrow\downarrow}_{k_1,k_2,\bar{s}}$ (shown as

squares) of each patch subject to the four constraints listed at the beginning of this section. From the boundary curve property of Bézier coefficients (property 2 of Section 5.3), the boundary coefficients of adjacent patches must be equal to ensure positional continuity. Constraint 1 implies that the coefficients around the boundary of N must be zero (zero coefficients are denoted by black circles). Constraint 2 above, together with the directional derivative property of Bézier coefficients (property 3 of Section 5.3), implies that the coefficients adjacent to the boundary coefficients must also be zero.

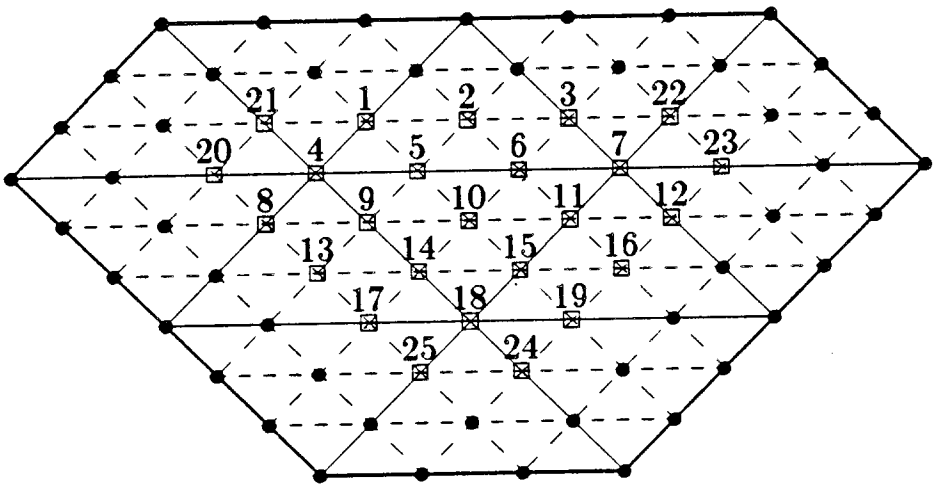

Figure 5.11. *A graph of N viewed from above the (u,v) plane. Black circles denote Bézier coefficients that must be zero. Squares denote Bézier coefficients that must be solved for.*

The Bézier reduction technique has allowed the system of unknowns to be reduced to 25 (possibly non-zero) coefficients $a_1, ..., a_{25}$, labeled $1, ..., 25$ in Figure 5.11. The remaining coefficients are determined by the equations implied by Constraint 3. We now examine the form of these equations.

Consider two adjacent basis patches shown in Figure 5.12(a). Derivatives along the boundary are already guaranteed to agree since the patches are constrained to have identically parametrized boundary curves. For C^1 continuity, we additionally require matching of the cross boundary derivative shown in Fig-

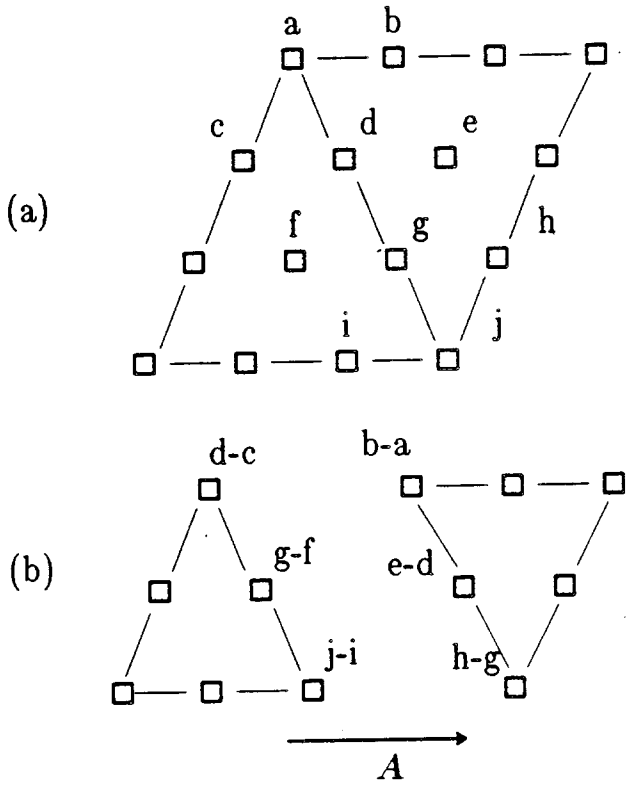

Figure 5.12. *Figure (a) shows two adjacent basis patches whose coefficients near the boundary are labeled a through j. Figure (b) shows the derivative surfaces corresponding to derivatives in the cross boundary direction indicated by* $\vec{A}$.

ure 5.12(b). The indicated cross boundary derivatives are equal if and only if

$$\begin{aligned} d - c &= b - a \\ g - f &= e - d \\ j - i &= h - g. \end{aligned} \tag{5.13}$$

However, for G^1 continuity, a simplified form of the bivariate Beta constraints (from Section 2.5.3), allows these conditions to be relaxed to

$$\begin{aligned} d - c &= \beta_1(b - a) \\ g - f &= \beta_1(e - d) \\ j - i &= \beta_1(h - g) \end{aligned} \tag{5.14}$$

for an arbitrary real number $\beta_1 > 0$.

Remark 5.1: *When more general Beta constraints were attempted, an inconsistent set of equations resulted. It may be that bivariate cubic polynomials simply do not have enough flexibility to allow more general constraints.*

The cross boundary derivative requirements (equations (5.13) or (5.14)) contribute three equations per interior boundary, and since there are 15 interior boundaries, it would seem that there are 45 equations for the 25 unknowns. However, the system of 45 equations contains only 24 linearly independent equations, leaving one degree of freedom for normalization of the overall height of the blending function N.

The normalization is chosen so that the spline surface $\mathbf{S}(u,v)$ is independent of the coordinate system in which the control vertices are specified. For this to occur, the blending function N must form a partition of unity in the sense that

$$\sum_{i_1,i_2} N(u-i_1, v-i_2) = 1. \tag{5.15}$$

Equation (5.15) can be interpreted as placing a graph of N at each lattice point in the parameter plane, requiring that the summation of the graphs be a constant one. To examine this further, consider the partial sum of two of the terms, the terms corresponding to $(i_1,i_2) = (0,0)$ and $(1,0)$, as shown in Figure 5.10. Note that the basis patch labeled $n^{\downarrow}_{0,0}$ with respect to $N(u,v)$ is superimposed on the basis patch labeled $n^{\downarrow}_{1,0}$ with respect to $N(u-1,v)$. Thus, the partial sum surface over darkened triangle is the sum of the basis patches $n^{\downarrow}_{0,0}$ and $n^{\downarrow}_{1,0}$. This partial sum surface can be computed by adding the polynomials in a straightforward way, or, as a consequence of the Bézier representation of the basis patches, the control net for the partial sum surface can be computed by adding the control nets for the basis patches. The complete summed surface over this triangle is the sum of all down basis patches; thus, the control net for the complete summed surface is the sum of the control nets of the all of the down patches of N. As a consequence of the convex hull property and linear independence of the Bézier basis functions, the sum surface will be unity if and only if its Bézier coefficients are identically one. Thus, the sum of the control nets of the down patches must result in a planar control net one unit above the (u,v) plane. That is,

$$\sum_{k_1,k_2 \text{ for down patches}} f^{\downarrow}_{k_1,k_2,\vec{i}} = 1 \tag{5.16}$$

for all $\vec{i}$ such that $|\vec{i}| = 3$. Recall that we only have one degree of freedom left to specify normalization; the constraint we choose is $\vec{i} = (1,1,1)$ for down patches.

For this case, every coefficient in equation (5.16) is zero, except for $f^{\downarrow}_{0,0,(1,1,1)}$, which is equal to a_{10} (see Figure 5.11). Thus, our normalization condition is $a_{10} = 1$.

When G^1 constraints similar to equations (5.14) were used along the $u_1 = 0$, $u_2 = 0$, and $u_3 = 0$ edges, an inconsistent set of equations resulted. The only constraints for which a consistent set could be constructed required C^1 continuity along two of the edges, and G^1 continuity along the third. For the case where C^1 constraints similar to equation (5.13) are used along the $u_2 = 0$ and $u_3 = 0$ boundaries, and the G^1 constraints from equation (5.14) are used along the $u_1 = 0$ boundaries, the unknowns are found to be

$$\begin{aligned}
a_1 &= a_2 = a_3 = a_7 = a_{12} = \frac{1}{\delta} \\
a_4 &= a_8 = a_{13} = a_{16} = a_{17} = a_{18} = \frac{\beta_1}{\delta} \\
a_5 &= a_6 = a_{11} = \frac{\beta_1 + 1}{\delta} \\
a_9 &= a_{14} = a_{15} = \frac{2\beta_1}{\delta} \\
a_{10} &= 1 \\
a_{20} &= a_{21} = a_{22} = a_{23} = a_{24} = a_{25} = 0
\end{aligned} \tag{5.17}$$

where $\delta = 2\beta_1 + 1$. The basis patches are now completely determined (see Figure 5.13). For example, the basis patch $n^{\downarrow}_{1,0}(\vec{u})$ has Bézier coefficients

$$f^{\downarrow}_{1,0,\vec{i}} = \begin{cases} a_4 = \frac{\beta_1}{\delta} & \text{for } \vec{i} = (3,0,0) \\ a_8 = \frac{\beta_1}{\delta} & \text{for } \vec{i} = (2,0,1) \\ 0 & \text{otherwise.} \end{cases} \tag{5.18}$$

One can verify that the normalization conditions (5.16) are satisfied by solution (5.17). A similar condition on the up patches is also satisfied. One can also verify that when $\beta_1 = 1$, the triangular cubic Beta-spline blending function reduces to the triangular cubic B-spline blending function. Thus, when $\beta_1 = 1$, a triangular cubic Beta-spline surface reduces to the triangular cubic B-spline surface defined by the same control net.

Remark 5.2: *Because of the Bézier reduction method of derivation, it is easy to see that the triangular cubic Beta-spline surface* $\mathbf{S}(u, v)$ *will lie in the convex hull of its control net. This follows from the fact that the blending functions form a partition of unity, and all the Bézier coefficients* $(a_1, ..., a_{25})$ *are non-negative,*

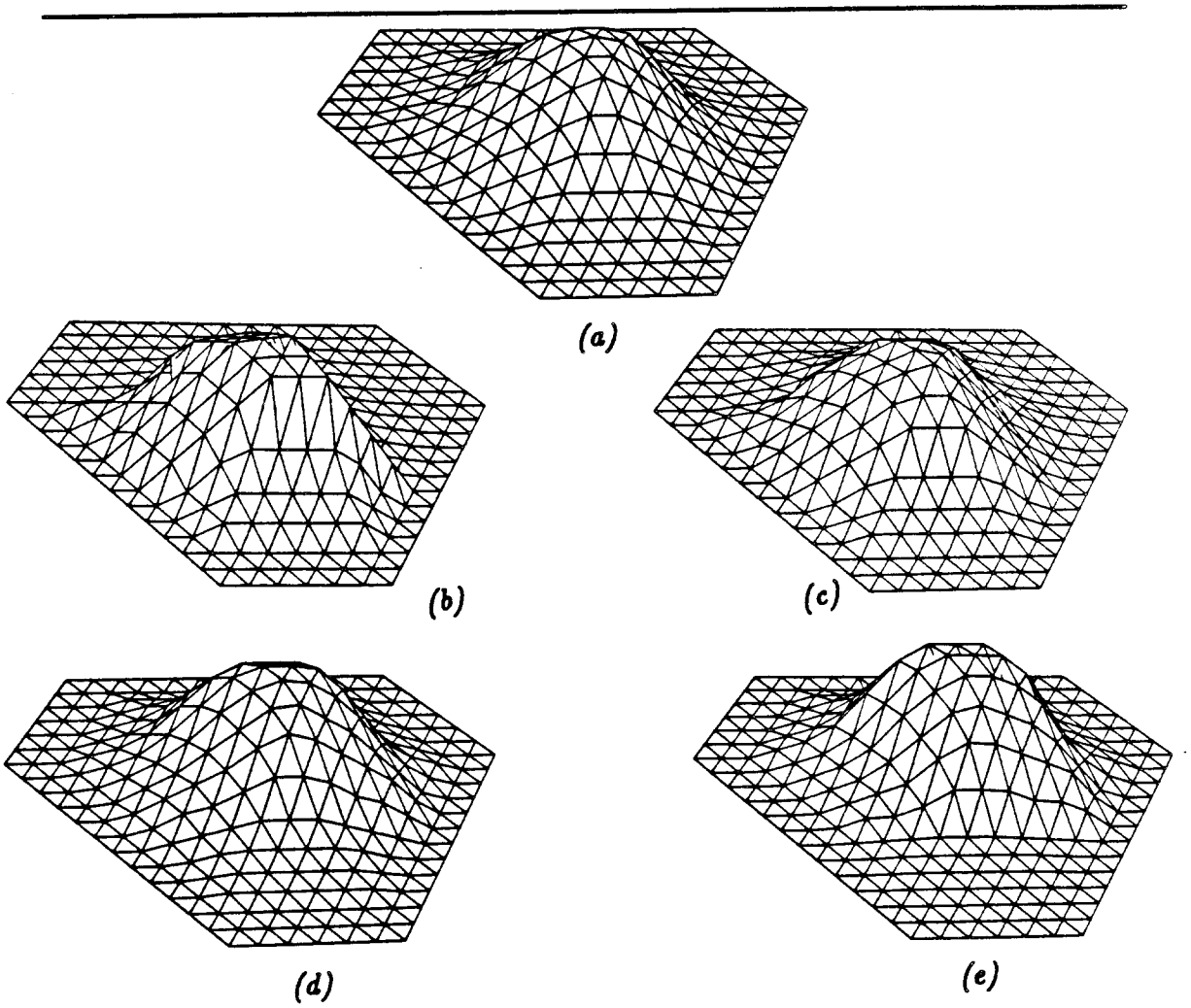

Figure 5.13. *Figure (a) is a plot of the triangular cubic Beta-spline with $\beta_1 = 1$, and is equivalent to the triangular cubic B-spline blending function. Figures (b), (c), (d), and (e) have β_1 set to .1, .5, 2, and 10, respectively. Reciprocal values were chosen to demonstrate the asymmetric behavior of β_1.*

implying that the blending function $N(u, v)$ is non-negative. Another advantage of the Bézier reduction method is the ease with which an evaluation algorithm can be developed. Such an algorithm is constructed in the next section. Thus, Bézier reduction has allowed the system of equations to be greatly reduced, generated an easy proof of the convex hull property, and provided the key to an efficient evaluation algorithm.

5.5.2. Evaluation Algorithm

Given a triangular control net $\mathbf{V}_{i_1,i_2}$ and a value of β_1, the triangular cubic Beta-spline can be calculated according to equation (5.11) using the basis patches derived in Section 5.5.1. Equation (5.11) defines the spline surface as a piecewise function $\mathbf{S}(u,v)$, for u and v varying over some portion of the parameter plane. We can also characterize the spline surface in terms of the triangular patches that comprise $\mathbf{S}$, in much the same way that a spline curve can be characterized by its curve segments (see Section 3.1). To do this, let $\mathbf{V}_{\vec{g}}$, $\vec{g} = (g_1, g_2)$, be a control vertex not "near" the boundary of the control net (this allows us to ignore boundary conditions on the surface). Associate two triangular surfaces patches with $\mathbf{V}_{\vec{g}}$, one an "up" patch $\mathbf{s}^{\uparrow}_{\vec{g}}(\vec{u})$, the other a "down" patch $\mathbf{s}^{\downarrow}_{\vec{g}}(\vec{u})$ (see Figure 5.14).

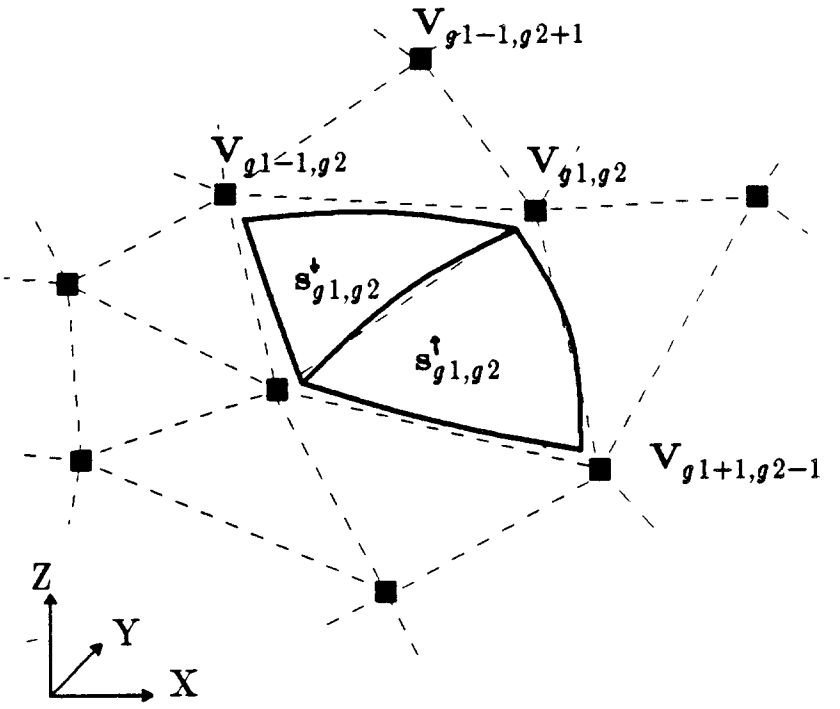

Figure 5.14. *The labeling of the surface patches associated with a control vertex not near the boundary of the control net. Dotted lines correspond to edges of the control net. Solid lines indicate patch boundaries on the spline surface.*

For convenience, let $\vec{k} = (k_1, k_2)$, $k_1, k_2 \in \{-1, 0, 1\}$. Using the labeling on the basis patches of $N(u, v)$ from Figure 5.9, it is not hard to show that

$$\mathbf{s}_{\vec{g}}^{\downarrow}(\vec{u}) = \sum_{|\vec{k}|=-1}^{1} \mathbf{V}_{\vec{g}+\vec{k}}\, n_{\vec{k}}^{\downarrow}(\vec{u}) \tag{5.19}$$

and

$$\mathbf{s}_{\vec{g}}^{\uparrow}(\vec{u}) = \sum_{|\vec{k}|=0}^{2} \mathbf{V}_{\vec{g}+\vec{k}}\, n_{\vec{k}}^{\uparrow}(\vec{u}) \tag{5.20}$$

where $\vec{u}$ refers to barycentric coordinates relative to the domain triangle of the basis patch. These expressions can be directly evaluated in the obvious way by evaluating the basis patches given their Bézier representation from equation (5.12). However, a more flexible method of evaluation based on *recursive subdivision* [7,18,46] is also possible.

The basic step of recursive subdivision algorithms is the "splitting" of the surfaces into smaller sub-parts. This process is continued recursively until the sub-parts are close to planarity, at which time they are approximated by polygons. Thus, recursive subdivision is a method of computing piecewise planar approximations to surfaces. Triangular cubic Beta-splines can be approximated on a patch by patch basis by converting the Beta-spline control vertices influencing the patch into Bézier control vertices that describe the same patch. The resulting Bézier control net can then be subdivided using Goldman's simplex subdivision algorithms [36].

The Bézier reduction method of derivation makes the conversion of Beta-spline control vertices into Bézier control vertices easy. Let $\mathbf{s}_{\vec{g}}^{\uparrow\downarrow}(\vec{u})$ be a patch of the spline surface given by (5.19) or (5.20). Substitute into these equations the form for the basis patches from equation (5.12) to yield

$$\begin{aligned}
\mathbf{s}_{\vec{g}}^{\uparrow\downarrow}(\vec{u}) &= \sum_{|\vec{k}|} \mathbf{V}_{\vec{g}+\vec{k}} \sum_{|\vec{i}|} f_{\vec{k},\vec{i}}\, b_{\vec{i}}^{3}(\vec{u}) \\
&= \sum_{\vec{i}} \left(\sum_{\vec{k}} \mathbf{V}_{\vec{k}+\vec{g}} f_{\vec{k},\vec{i}}^{\uparrow\downarrow} \right) b_{\vec{i}}^{3}(\vec{u}) \\
&= \sum_{\vec{i}} \mathbf{W}_{\vec{g},\vec{i}}^{\uparrow\downarrow} b_{\vec{i}}^{3}(\vec{u})
\end{aligned} \tag{5.21}$$

where

$$\mathbf{W}_{\vec{g},\vec{i}}^{\uparrow\downarrow} = \sum_{\vec{k}} \mathbf{V}_{\vec{k}+\vec{g}} f_{\vec{k},\vec{i}}^{\uparrow\downarrow} \tag{5.22}$$

are the Bézier control vertices that we seek. These vertices can be input to a recursive simplex subdivision algorithm to produce a polygonal (actually triangular) approximation to the original Beta-spline surface. Figure 5.15 shows a sequence of images produced in this way.

Figure 5.15 also shows the effect of shape parameter modification. Note that as the shape parameter is increased, the patches deform, but always in a way that preserves tangent plane continuity between the patches. The annoying undulations in the left boundary of the spline surface seem to be an artifact of the "bias-like" nature of the shape parameter. That is, the shape parameter tends to skew the blending function, and hence, the spline surface. It may be possible to counteract the boundary undulations through a judicious choice of boundary conditions, perhaps by placing multiple vertices at the boundaries, or by adding *phantom vertices* [3,4,8] to "straighten out" the boundary curves.

5.6. Summary

In this chapter, some examples of the use of geometric continuity for spline surfaces composed of triangular patches were examined. These included a discussion of Bézier triangles, and the construction of a new surface technique called the triangular cubic Beta-spline.

The triangular cubic Beta-spline is a geometrically continuous analog of the triangular cubic B-spline. It is a G^1 continuous spline technique possessing one global shape parameter. For positive values of the shape parameter, the convex hull property is exhibited. When the shape parameter is set to the "default" value of one, the triangular cubic Beta-spline reduces to the triangular cubic B-spline.

An evaluation algorithm for triangular cubic Beta-splines was developed, and examples of the effect of changing a shape parameter were given. These examples show that as the shape parameter is increased, undesirable undulations in the spline boundaries develop. Unless the undulations can be counteracted in some way, the triangular cubic Beta-spline will probably not be useful in a practical setting.

The triangular cubic Beta-spline is still interesting, at least in a theoretic sense, because it establishes the existence of geometrically continuous triangular surfaces related to the triangular B-splines. Thus, it may be possible to construct the geometrically continuous analog of the triangular quartic B-spline. This technique, call it the *triangular quartic Beta-spline*, would be a G^2 representation

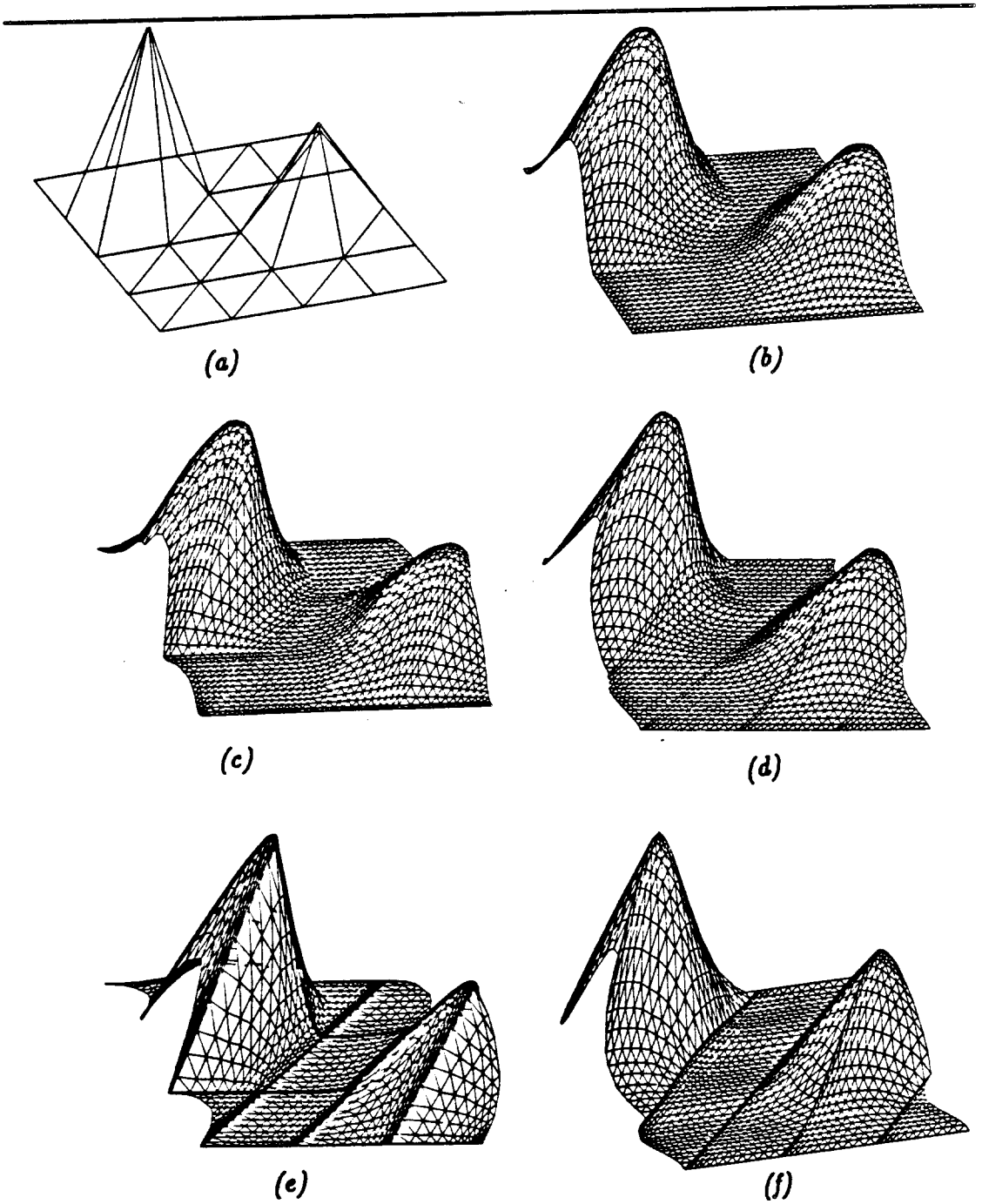

Figure 5.15. *The triangular surfaces above are all defined by the same control net shown in Figure (a); they differ only in the value of* β_1. *Figure (b) has* $\beta_1 = 1$, *and Figures (c), (d), (e), and (f) have* β_1 *set to* $.5, 2, .1,$ *and* 10, *respectively. They were rendered using the recursive simplex subdivision algorithm described in Section 5.5.2.*

supported over 24 triangular patches. In the case of Beta-spline curves, the first order shape parameter behaves asymmetrically, as it does in the triangular cubic Beta-spline, but the second order parameter β_2 behaves symmetrically. It is hoped that the same trends carry over to triangular surfaces, meaning that the second order parameters would behave symmetrically. The investigation of these questions is a topic of current research.

As suggested above, there are numerous topics for continued research, including:

- An understanding of why only a special case of the Beta constraints could be used in the construction of the triangular cubic Beta-spline.
- The introduction of a local shape parameter.
- The study of boundary conditions to remove the undulations in the spline boundaries.
- The construction of the triangular quartic Beta-spline, and the possible symmetric behavior of the second order shape parameters.
- A general evaluation algorithm (if one exists) for triangular Beta-splines of arbitrary order.

6

Foundations of Geometric Continuity

In Chapter 2, geometric continuity was characterized by requiring the existence of GO-equivalent parametrizations that meet with parametric continuity. Although this approach is conceptually simple, it is difficult to prove some of the statements that were presented as plausible in Chapter 2. In this chapter, we present a formalism that is better suited to proving statements concerning geometric continuity. Our formalism is based on the theory of differentiable manifolds.

We begin by motivating the use of manifold theory. Sections 6.2, 6.3, and 6.4 introduce tools from topology, advanced calculus, and manifold theory that are useful for developing geometric continuity. In Sections 6.5, 6.6, 6.7, and 6.8, the spline construction problem is described in the framework of manifolds and geometric continuity is defined. In Section 6.9, the Beta constraints of arbitrary parametric dimension are derived and shown to be necessary and sufficient for geometric continuity. Finally, in Section 6.10, several equivalent definitions of geometric continuity are established.

6.1. Introduction

Splines as piecewise differentiable functions have their roots in approximation theory. where much of the work has been focused on the use of splines to approximate real-valued functions. Since functions are being approximated, this is an inherently non-parametric application. B-splines, originally discussed by Curry & Schoenberg [21], were found to be exceptionally useful in this context. The discovery by Cox [20], and independently by deBoor [11], of a stable

evaluation algorithm for B-splines sparked a great deal of interest in the use of parametric B-splines in CAGD [39,51]. However, the generalization of B-spline curves into the parametric realm had an interested and very subtle side-effect. Parametric B-splines curves are piecewise parametric functions, but the individual parametric curve segments share the same parameter space — the real line. Similarly, parametric tensor product B-spline surfaces can be viewed as piecewise surfaces, but again, the surface patches are defined on the same parameter space — the plane. These preconceptions naturally lead to a development of parametric continuity.

Since parametric functions are of interest in CAGD, it is possible, and as we will see, desirable, to allow each curve segment or surface patch to be defined on its own, distinct parametric domain. The central theme of this chapter, and indeed, of this work as a whole, is to understand how smooth parametric splines can be constructed with the basic premise that each curve segment or surface patch is defined on its own domain. We are therefore proposing an inherently parametric view.

In the inherently parametric view of spline construction, one begins with a collection of parametrizations defined on initially unrelated and disjoint domains, with the goal of stitching the parametrizations together to form a smooth, composite image. Since there can be any number of parametrizations fitting together in arbitrarily complex patterns, it is difficult to see how to make sense of such an unstructured situation. The method espoused here is to introduce a *differentiable manifold* as a central platform upon which the parametric domains can be related.

Intuitively, differentiable manifolds are smooth, continuous sets of points upon which calculus can be performed. Manifolds are particularly attractive for our purposes because their mathematical structure makes it very easy to make *coordinate independent statements*. In Section 6.4.2, we show that coordinate independent statements are equivalent to parametrization independent statements in the construction of splines. Since we seek a parametrization independent measure of continuity, the coordinate independent properties of manifolds are therefore very convenient. The use of manifold theory also makes it easy to describe geometric continuity for splines of arbitrary parametric dimension, thereby unifying the development for curves, surfaces, volumes, etc. Manifold theory also allows spline surfaces to be defined on domains other than the parameter plane. In fact, any differentiable manifold can be used as a domain.

There may also be secondary benefits from using a formalism based on manifold theory. To describe geometric continuity using manifold theory requires the establishment of a firm connection between splines in CAGD and manifold theory. The connection is established by providing a rigorous definition of a *parametric spline*, stated in manifold theoretic terms. Having done this, many results in the field of manifolds may have direct application in CAGD.

The major disadvantage of a formalism based on manifold theory is a pragmatic one. Those wishing to understand geometric continuity at a fundamental level must either be familiar with manifolds, or be willing to spend the time to become familiar with them. However, it is important to point out that in the "old language" of Chapter 2, parametric continuity seemed natural, and geometric continuity was developed as a rather subtle *generalization*. On the other hand, in the "new language" of manifolds, geometric continuity is natural, and parametric continuity is developed as a rather subtle *special case*. For this reason, if for no other, we feel that the burden of introducing manifold theory is justified.

As a high-level road map of the material to come, the development is based on casting the spline construction problem into the language of manifolds. The idea is to start with a collection of parametrizations comprising a parametric spline, where each parametrization is defined on its own domain. The parametrizations and their domains are then "lifted" onto an infinitely differentiable (C^∞) manifold $\mathcal{P}$ to obtain a coordinate free, and hence a parametrization free, characterization of the spline known as an *abstract spline*. The abstract spline is viewed as a map from the manifold $\mathcal{P}$ into $\Re^m$, which we identify with Euclidean m-space. Actually, the maps could be into any manifold with suitable dimension, and the manifold used as a domain needs to be only as differentiable as the spline that is to be constructed on it. We make the above restrictions only for concreteness and clarity; the relaxation of these restrictions poses little technical difficulty. In fact, splines into manifolds other than Euclidean space have application in animation control, as recently shown by Gabriel & Kajiya [35] and Shoemake [56].

By requiring smoothness of the abstract spline, a parametrization independent measure of continuity is achieved in a natural way. The characterization of geometric continuity then becomes a problem of determining how smoothness of the abstract spline determines continuity conditions on the parametrizations.

As a word of warning, it is not the purpose of this chapter to teach the reader manifold theory. Rather, we present a brief introduction to the central results of the theory that have direct application to geometric continuity. We begin with

some preliminary definitions and results from topology, advanced calculus, and differentiable manifold theory. The preliminary material is not intended to be precise, only informative. More complete introductions to manifold theory can be found in Boothby [14], Brickell [15], and Spivak [57,58].

6.2. Some Concepts from Elementary Topology

The concept of an open set is of primary importance in topology, and since manifolds assume an underlying topological structure, it is natural that opens sets play an important role in our development. While it is possible to define open sets for spaces other than Euclidean space, such complications are not necessary for our purposes. Instead, we follow the development of open and closed subsets of Euclidean space put forth by Spivak [57].

The *open interval* (a,b), $a,b \in \Re$, with $a < b$, is the prototypical open set from which all other open sets will be defined. Let $(a_1,b_1),...,(a_m,b_m)$ be m prototypical open intervals. The set $(a_1,b_1) \times \cdots \times (a_m,b_m)$ is called an *open rectangle* in $\Re^m$. More generally, a set $U \in \Re^m$ is called *open* if for each $x \in U$, there is an open rectangle A such that $x \in A \subset U$. A subset C of $\Re^m$ is called *closed* if its complement in $\Re^m$ (written $\Re^m - C$) is open.

A *homeomorphism* is a continuous, 1-1, onto map whose inverse is also continuous. Intuitively, a homeomorphism is an *elastic* map in which neighboring points get mapped to neighboring points. If f is a homeomorphism from a set S_1 onto a set S_2 (written $f : S_1 \rightarrow S_2$), then S_1 and S_2 are said to be *homeomorphic.*

A *ball* in $\Re^m$ *of radius* ϵ *about a point* $q \in \Re^m$, denoted $\mathbf{B}_\epsilon^m(q)$, is the set of points in $\Re^m$ whose distance from q is strictly less than ϵ. Thus, a ball in $\Re^1$ is an open interval, a ball in $\Re^2$ is the interior of a circle, a ball in $\Re^3$ is the interior of a sphere, and so on. The *closure of* $\mathbf{B}_\epsilon^m(q)$, denoted $\bar{\mathbf{B}}_\epsilon^m(q)$, is the set of points in $\Re^m$ whose distance from q is less than or equal to ϵ. One can show that balls are open sets, and that closures of balls are closed sets.

A set A of $\Re^m$ is said to be *bounded* if it is contained in a ball of finite radius. Finally, a set A of $\Re^m$ is said to be *compact* if it is closed and bounded.*

6.3. A Brief Review of Multivariate Calculus

* This narrow definition of compactness suffices for our purposes; for a more general definition, see Spivak [57].

Definition 6.1: *Let f be a map into $\Re^m$ defined on an open set U of $\Re^n$, i.e., $f : U \to \Re^m$. The restriction of f to $A \subset U$ is written as $f|A$. f is said to be of class C^r at $x \in U$ if all partial derivatives up to order r exist and are continuous at x. f is said to be of class C^r if it is C^r at all points $x \in U$. If f is C^r for all r, then we say that f is C^∞.*

Let $f : U \to \Re^m$ be a C^r map, U an open set in $\Re^n$, $r \geq 1$, and let $(x_1, ..., x_n)$ and $(y_1, ..., y_m)$ be coordinate systems on $\Re^n$ and $\Re^m$, respectively. We extend the notation of Section 2.5 and use $f^{\vec{k}}$, $\vec{k} = (k_1, ..., k_n)$, to mean

$$f^{\vec{k}} \equiv \left(\frac{\partial^{|\vec{k}|} f}{\partial {x_1}^{k_1} \cdots \partial {x_n}^{k_n}} \right). \tag{6.1}$$

Let f_i denote the i^{th} coordinate function of f relative to $(x_1, ..., x_n)$ and $(y_1, ..., y_m)$, and let $Df(x)$ denote the Jacobian matrix of f at $x \in U$:

$$Df(x) \equiv \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{pmatrix}. \tag{6.2}$$

If f is a complicated functional expression, we will sometimes write $D[f](x)$ instead of $Df(x)$. It is sometimes convenient to express equation (6.2) more compactly by writing the s^{th} column as $f^{\vec{j}_s}(x)$, $\vec{j}_s$ as defined in Section 1.2. With this notation, equation (6.2) becomes

$$Df(x) \equiv [f^{\vec{j}_1}(x) \cdots f^{\vec{j}_n}(x)]. \tag{6.3}$$

Let f, n, and m be as above with $n \leq m$. The rank of f at $x \in U$, denoted $\text{rank}(f(x))$, is defined to be the rank of $Df(x)$. f is said to be regular at $x \in U$ if $\text{rank}(f(x)) = n$, that is, if f is of full rank; f is regular if it is regular at all points $x \in U$. In the special case where $n = m$, the Jacobian of f at x, denoted by $Jf(x)$, is defined to be the determinant of $Df(x)$.

Remark 6.1: *Let $f : U \to \Re^n$ be a C^r map, U an open set in $\Re^n$. The statement that f is regular at $x \in U$ is equivalent to the condition $Jf(x) \neq 0$.*

Theorem 6.1: (The Chain Rule) *Let U be an open set in $\Re^n$ and V be an open set in $\Re^m$. Let $f : U \to V$ and $g : V \to \Re^p$ be C^r at $x \in U$ and $f(x) \in V$, respectively. The composite map $h = g \circ f : U \to \Re^p$ is C^r at $x \in U$ and its Jacobian matrix is given by*

$$Dh(x) = Dg(f(x)) \cdot Df(x) \tag{6.4}$$

where $\cdot$ denotes matrix multiplication.

Proof: c.f. Boothby [14], Theorem 2.3 and Corollary 2.4. ∎

Theorem 6.2: *Let f, g, h be as above where $n = m = p$. If f and g are regular and C^r at $x \in U$ and $f(x) \in V$, respectively, then $h = g \circ f$ is regular and C^r at $x \in U$.*

Proof: Theorem 6.1 shows that h is C^r at x; regularity follows from Remark 6.1 and Theorem 6.1. ∎

In Section 6.2, the notion of a homeomorphism was introduced. Although a homeomorphism f must be continuous, it need not be differentiable. However, if f and its inverse f^{-1} are r-times differentiable, then f is said to be a *C^r differentiable homeomorphism*, more commonly called a *C^r-diffeomorphism*. A C^∞-diffeomorphism is often called simply a *diffeomorphism*. If f is a diffeomorphism from a set S_1 onto a set S_2, then S_1 and S_2 are said to be *diffeomorphic*. The next theorem, known as the *Inverse Function Theorem*, shows that regular C^r maps are locally C^r-diffeomorphisms.

Theorem 6.3: (The Inverse Function Theorem) *Let $f : U \to \Re^n$ be a regular C^r map, $r = 1, 2, ..., \infty$, U an open subset of $\Re^n$. At each point in $f(U)$, a local inverse map f^{-1} exists and is regular and C^r; hence, f is locally a C^r-diffeomorphism. Moreover, if $x \in U$ and $y = f(x)$, then the Jacobian matrix of f^{-1} at y is given by*

$$Df^{-1}(y) = (Df(x))^{-1} \tag{6.5}$$

where the inverse on the right side refers to the inverse of the matrix $Df(x)$.

Proof: The Inverse Function Theorem is proved in most standard texts of advanced calculus, differential geometry, and differentiable manifolds (c.f. Boothby [14], Chapter 2). ∎

Remark 6.2: *The Inverse Function Theorem guarantees that regular maps are locally invertible, but it in no way implies that they are globally invertible. For instance, the curve depicted in Figure 2.3a can be generated by a regular univariate map, but due to the cross over point, there does not exist a global inverse map.*

The following theorem is a slight generalization of the Inverse Function Theorem that will prove to be useful in Section 6.9

Theorem 6.4: *Let U be an open set of $\Re^n$, and let $V \subset U$ be compact. Let $f : U \to \Re^n$ be 1-1, regular, and C^∞ on V. Then there exists an open neighborhood W of V in U such that f is a diffeomorphism on W.*

Proof: For a sketch of a proof of this theorem, see Guillemin [41], exercise 10, page 19. ∎

Notation: Let $f, g : U \to \Re^m$ be two C^∞ maps, U an open set of $\Re^n$, and let x be a point in U. We use the notation

$$(f)_x \overset{C^r}{=} (g)_x \tag{6.6}$$

to mean

$$f^{\vec{k}}(x) = g^{\vec{k}}(x) \tag{6.7}$$

for all $\vec{k} = (k_1, ..., k_n)$ such that $|\vec{k}| \leq r$. For example, let $n = 2$, $m = 3$, let (x_1, x_2) be a coordinate system on $\Re^2$, and let f_1, f_2, f_3 and g_1, g_2, g_3 be the component functions of f and g relative to a coordinate system on $\Re^3$. The expression

$$(f)_x \overset{C^2}{=} (g)_x \tag{6.8}$$

is shorthand for the 18 scalar conditions

$$\begin{aligned}
f_i(x) &= g_i(x) && (\vec{k} = (0,0)) \\
\frac{\partial f_i}{\partial x_1} &= \frac{\partial g_i}{\partial x_1} && (\vec{k} = (1,0)) \\
\frac{\partial f_i}{\partial x_2} &= \frac{\partial g_i}{\partial x_2} && (\vec{k} = (0,1)) \\
\frac{\partial^2 f_i(x)}{\partial^2 x_1} &= \frac{\partial^2 g_i(x)}{\partial^2 x_1} && (\vec{k} = (2,0)) \\
\frac{\partial^2 f_i(x)}{\partial x_1 \partial x_2} &= \frac{\partial^2 g_i(x)}{\partial x_1 \partial x_2} && (\vec{k} = (1,1)) \\
\frac{\partial^2 f_i(x)}{\partial^2 x_2} &= \frac{\partial^2 g_i(x)}{\partial^2 x_2} && (\vec{k} = (0,2))
\end{aligned}$$

for $i = 1, 2, 3$.

6.4. Elementary Manifold Theory

For our purposes, it will be sufficient to treat a *manifold* $\mathcal{M}$ of *dimension* m, more commonly called an *m-manifold*, as a "continuous" set of points that is locally Euclidean.* That is, for every point q of $\mathcal{M}$, there is a neighborhood U of q that is homeomorphic to an open set U' in $\Re^m$.

* For a more precise definition, the reader is referred to one of the standard texts mentioned in Section 6.1.

Example 6.1: *Any open set in $\Re^m$ is automatically an m-manifold, the homeomorphism being the identity map. A more interesting example of a manifold is the two-sphere, denoted S^2. It consists of all points in $\Re^3$ that lie a unit distance from the origin. One proves that S^2 is a 2-manifold by demonstrating that for every point q of S^2, a homeomorphism from a neighborhood U of q into $\Re^2$ exists. One way to do this is to choose U small enough that points of U can be projected in a 1-1 fashion onto the tangent plane of the sphere at q (see Figure 6.1). The projection is a homeomorphism of U onto an open set in $\Re^2$ (the tangent plane).*

•

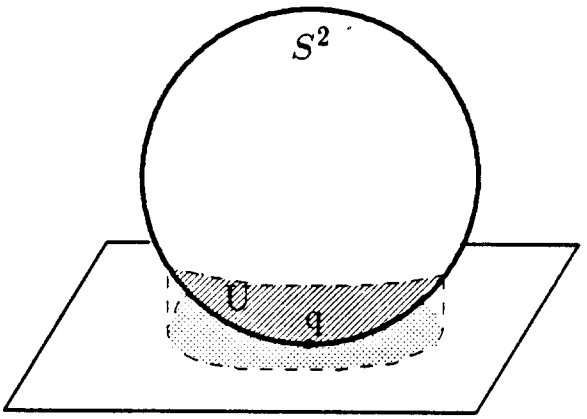

Figure 6.1. *The two-sphere can be shown to be a 2-manifold by projecting neighborhoods of points onto the tangent plane.*

Although S^2 is a 2-manifold, and is therefore locally Euclidean, it is not globally homeomorphic to $\Re^2$. Thus, it is not possible to introduce a single coordinate system on S^2 that is non-degenerate. This motivates the introduction of local coordinate systems. We can assign coordinates to a point q of an m-manifold as follows. Let U be a neighborhood of q that is homeomorphic to an open set U' of $\Re^m$, and let ϕ be the corresponding homeomorphism, called a *chart*. Assign to q the coordinates of x, where $x = \phi(q)$ (see Figure 6.2). In this way, the chart ϕ, together with its domain U, provides a *local coordinate system* on the manifold. We more commonly say that (U, ϕ) is a *coordinate neighborhood of q*.

It was mentioned above that an open subset of $\Re^m$ is automatically a manifold. However, such a manifold does not contain its boundary points. Keeping in mind

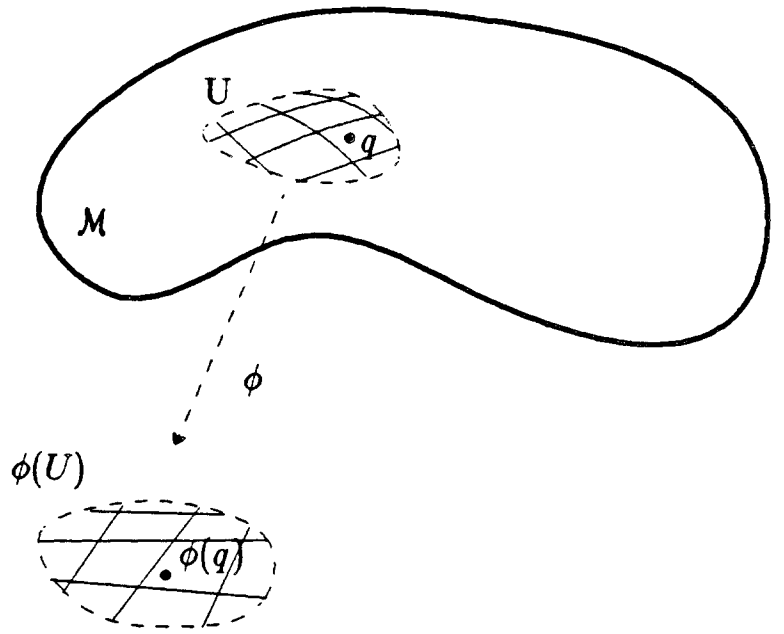

Figure 6.2. *Charts provide local coordinate systems on manifolds.*

that we are ultimately interested in casting the spline construction problem into the framework of manifolds, and noting that spline curves and surfaces usually do contain their boundary points, it is useful to define a manifold that contains its boundary. Let $\mathcal{H}^m$ denote the *positive half space of* $\mathcal{R}^m$, defined by

$$\mathcal{H}^m = \{(x_1, x_2, ..., x_m) \in \mathcal{R}^m | x_m \geq 0\}. \tag{6.9}$$

An *m-manifold with boundary* is a continuous set of points $\mathcal{M}$ such that each $q \in \mathcal{M}$ has a neighborhood that is homeomorphic to $\mathcal{R}^m$ or $\mathcal{H}^m$. If the neighborhood is homeomorphic to $\mathcal{R}^m$, q is called an *interior point* of the manifold, otherwise it is called a *boundary point.* The collection of boundary points, denoted by $\partial\mathcal{M}$, is called the *boundary of* $\mathcal{M}$. If $\mathcal{M}$ is an m-manifold with boundary, then $\partial\mathcal{M}$ is an $(m-1)$-manifold (cf. Boothby [14], Chapter 6).

The notions of differentiability and calculus are extended to manifolds by requiring a certain smoothness condition between overlapping charts. In particular, let (U, ϕ) and (V, ψ) be two overlapping coordinate neighborhoods, that is, their domains have a non-null intersection: $U \cap V \neq \emptyset$. The charts ϕ and ψ are said to be C^r*-compatible* if the map $\psi \circ \phi^{-1} : \phi(U \cap V) \rightarrow \psi(U \cap V)$ is C^r in the advanced calculus sense of Definition 6.1 (see Figure 6.3). If the coordinate neighborhoods

are C^r-compatible for every r, we say they are C^∞-compatible. We call the map $\psi \circ \phi^{-1}$ a *transition function* because it relates the coordinate system provided by ϕ to the coordinate system provided by ψ.

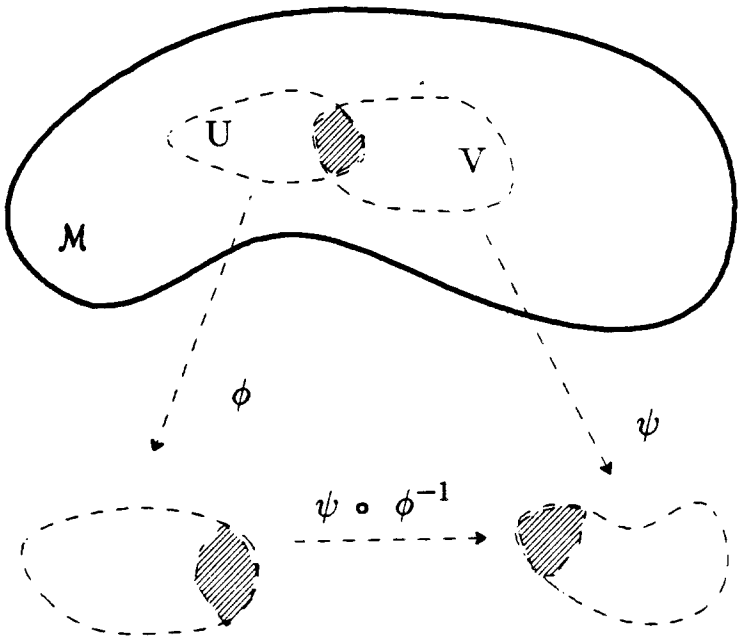

Figure 6.3. *Compatibility of overlapping charts.*

Definition 6.2: *A C^r-differentiable structure or atlas on a manifold M is a collection $\mathcal{A}$ of coordinate neighborhoods (U_α, ϕ_α) such that:*

(i) the U_α cover M.

(ii) if (U_α, ϕ_α) and (U_β, ϕ_β) are overlapping coordinate neighborhoods in $\mathcal{A}$, then they are C^r-compatible.

(iii) any coordinate neighborhood (W, ψ) compatible with elements of $\mathcal{A}$ is itself in $\mathcal{A}$. Thus, $\mathcal{A}$ is a maximal collection of compatible coordinate neighborhoods.

A C^r-differentiable manifold is a manifold together with a C^r differentiable structure. A C^∞-differentiable manifold is simply called a differentiable manifold. A C^r-differentiable manifold with boundary is a manifold with boundary together with a C^r-differentiable structure. A C^∞-differentiable manifold with boundary is simply called a differentiable manifold with boundary.

Strictly speaking, a differentiable manifold is a tuple $(M, \mathcal{A})$, M a manifold, and $\mathcal{A}$ a differentiable structure. However, we will almost always be lax and simply

refer to $\mathcal{M}$ as a differentiable manifold, only implicitly referring to the differentiable structure. While this practice will lead to no difficulty in our development, there may be future applications where the differentiable structure must be stated explicitly.

Common examples of differentiable manifolds are the unit interval $(0,1)$, the unit circle, the Euclidean plane, the two-sphere, the torus, etc. The closed interval $[0,1]$ and $\bar{\mathbf{B}}_1^2(0)$ are examples of differentiable manifolds with boundaries. There are also many examples of differentiable manifolds that are not defined as subsets of Euclidean space. For instance, the set of all lines in $\Re^m$ that contain the origin can be shown to be a manifold on which a differentiable structure can be defined. Similarly, it is possible to define a differentiable structure on the set of non-singular $m \times m$ matrices.

Remark 6.3: *A collection of coordinate neighborhoods satisfying only properties* (i) *and* (ii) *of Definition 6.2 is called a* C^r*-differentiable basis for* $\mathcal{M}$*. It can be shown that a differentiable basis for* $\mathcal{M}$ *uniquely determines a differentiable structure for* $\mathcal{M}$ *(cf. Boothby [14]).*

6.4.1. Orientable Manifolds

As mentioned in Chapter 2, it is often advantageous to endow a spline with an orientation. To do this in the manifold approach requires the introduction of the notion of an ***orientable manifold.***

The orientation of a vector space can be defined by considering two sets of basis vectors, along with the transformation matrix that relates them. The basis sets are said to be ***coherently oriented*** if the transformation matrix relating them has a positive determinant. Equivalently, the bases are said to have the ***same orientation*** and the transformation matrix is said to be ***orientation preserving.*** If the transformation matrix has a negative determinant, the bases are said to have ***opposite orientation,*** and the transformation matrix is said to be ***orientation reversing.*** Thus, given a vector space V, and a basis B for V, an orientation can be assigned to V. Namely, the orientation provided by B.

Example 6.2: *As an example of an oriented vector space, let* V *be Euclidean 3-space, and choose* B *to be a right handed orthonormal basis. All other bases for* V *that have the same orientation as* B *are also right handed. Thus,* V *and* B *together define a right handed Euclidean space.* •

The notion of orientation can be extended to manifolds as follows: Let $\mathcal{M}$ be a differentiable manifold, and let (U, ϕ) and (V, ψ) be two overlapping coordinate neighborhoods. These coordinate neighborhoods are said to be *coherently oriented* if the transition function $\phi \circ \psi^{-1}$ has an orientation preserving Jacobian matrix. The manifold $\mathcal{M}$ is said to be *orientable* if there exists an *oriented basis*; that is, a basis of coherently oriented coordinate neighborhoods. Euclidean space, the two-sphere, and the torus are examples of orientable manifolds. The Möbius strip and the Klein bottle are the most famous examples of non-orientable manifolds.

6.4.2. Maps on Manifolds

Having developed the underlying structure of a differentiable manifold, it is natural to study maps defined on them. In particular, we are interested in extending the notions of continuity, differentiability, and regularity to maps defined on manifolds.

Let $\mathcal{N}$ be a differentiable n-manifold, let W be an open set on $\mathcal{N}$, and let f be a map defined on W; that is, $f : W \to \Re^m$. The map f is an *abstract map* in the sense that its "action" does not depend on the coordinates assigned to points of W. For instance, the projection map that was used in Example 6.1 was completely characterized without the introduction of coordinates. However, if we are to do computations with f, it is desirable to express it in terms of local coordinates; this can be done as follows. Let (U, ϕ) be a coordinate neighborhood of $\mathcal{N}$ such that $W \cap U \neq \emptyset$, and let $x = (x_1, ..., x_n)$ denote coordinates assigned to points on $\mathcal{N}$ by (U, ϕ); that is, $\phi(q) = x = (x_1, ..., x_n)$, $q \in U$. f can be expressed relative to (U, ϕ) as a map $\hat{f}$ that assigns to each x an image point in $\Re^m$:

$$\hat{f}(x) = (\hat{f}_1(x), \cdots, \hat{f}_m(x)) \tag{6.10}$$

where $\hat{f}_1, ..., \hat{f}_m$ are the coordinate functions of $\hat{f}$. Equation (6.10) can be written more concisely in terms of the abstract map f as (see Figure 6.4)

$$\hat{f}(x) = f \circ \phi^{-1}(x), \qquad x \in \phi(W \cap U). \tag{6.11}$$

The map $\hat{f} = f \circ \phi^{-1}$ is said to be a *representation of f relative to (U, ϕ)*. Let $\tilde{f}$ be another representation of f relative to a coordinate neighborhood (V, ψ), where $W \cap U \cap V \neq \emptyset$, and let $y = (y_1, ..., y_n)$ denote coordinates assigned by (V, ψ); that is,

$$\tilde{f}(y) = f \circ \psi^{-1}(y), \qquad y \in \psi(W \cap V). \tag{6.12}$$

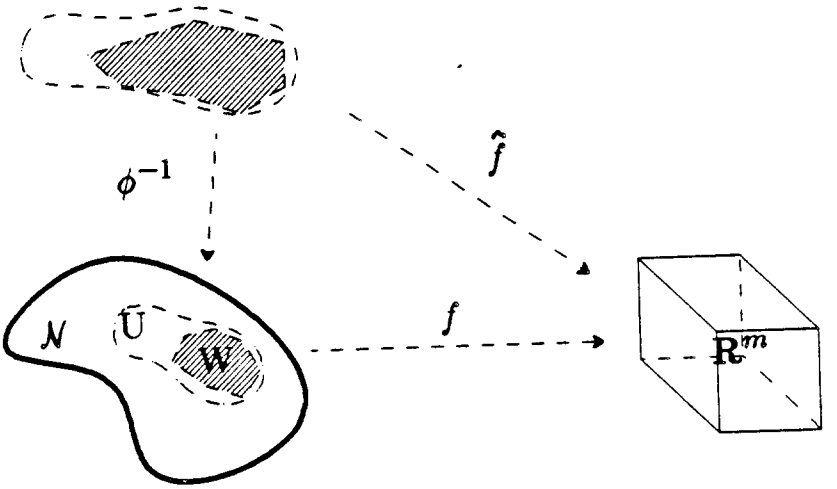

Figure 6.4. *The abstract map $f : W \to \Re^m$ has a representation $\hat{f} = f \circ \phi^{-1}$ relative to a coordinate neighborhood (U, ϕ).*

The two representations $\hat{f}$ and $\tilde{f}$ can be shown to be related on the open set $W \cap U \cap V$ by solving equation (6.11) for f, followed by substitution into equation (6.12) to yield

$$\tilde{f}(y) = \hat{f} \circ \phi \circ \psi^{-1}(y), \qquad y \in \psi(W \cap U \cap V). \tag{6.13}$$

Remark 6.4: *Equation (6.13) can be interpreted in two ways: As discussed above, $\hat{f}$ and $\tilde{f}$ are two different representations of the same abstract map f (see Figure 6.5). That is, $\hat{f}$ and $\tilde{f}$ represent the same action in two different local coordinate systems. In the terminology of Chapter 2, equation (6.13) states that $\hat{f}$ and $\tilde{f}$ are reparametrized versions of one another, with $\phi \circ \psi^{-1}$ playing the role of the change of parameter. Thus, when viewed in the language of manifolds, the process of reparametrization corresponds to a change of local coordinates in the representation of an abstract map.*

The observation made in Remark 6.4 is crucial since it implies that finding a definition of continuity that is parametrization independent is equivalent to finding a definition of continuity that is independent of the choice of local coordinates. Such a definition, which we now present, is the standard definition of differentiability of an abstract map on a manifold.

Definition 6.3: *Let $\mathcal{N}$ be a differentiable n-manifold. A map $f : W \to \Re^m$, W an open subset of $\mathcal{N}$, is said to be C^r at $q \in W$, $r = 1, 2, ..., \infty$, if there exists a*

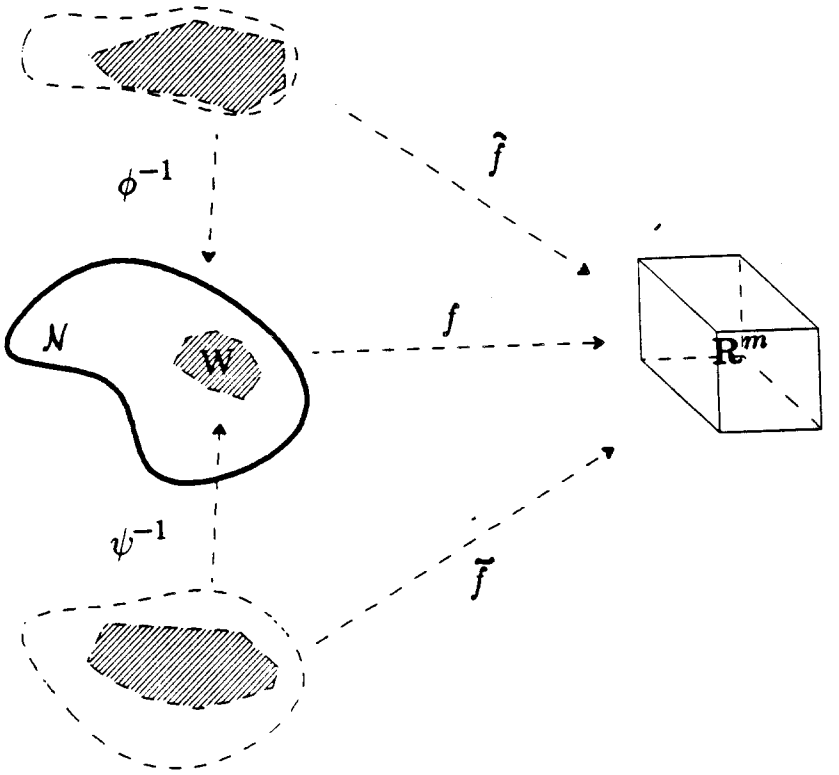

Figure 6.5. *$\hat{f}$ and $\tilde{f}$ are two representations for the same abstract map f*

coordinate neighborhood (U, ϕ) of q such that $\hat{f} = f \circ \phi^{-1}$ is C^r at $\phi(q)$ in the sense of Definition 6.1. f is C^r if it is C^r at each $q \in W$.

Remark 6.5: *Using the C^∞ compatibility of overlapping coordinate neighborhoods, it can be verified that the definition of differentiability for an abstract map is independent of the choice of local coordinate neighborhoods. That is, if (U, ϕ) and (V, ψ) are two coordinate neighborhoods of a point $q \in W$, and if $f : W \to \Re^m$, then f is C^r with respect to (U, ϕ) if and only if it is C^r with respect to (V, ψ). We therefore have:*

Claim 6.1: *Let $\mathcal{N}$ be a differentiable manifold, $q \in \mathcal{N}$, V a neighborhood of q, and ψ, ϕ two charts on V. If $f, g : V \to \Re^m$ are C^∞ maps, then for any r, the expression*

$$\left(f \circ \psi^{-1}\right)_{\psi(q)} \overset{C^r}{=} \left(g \circ \psi^{-1}\right)_{\psi(q)} \tag{6.14}$$

holds if and only if

$$\left(f \circ \phi^{-1}\right)_{\phi(q)} \overset{C^r}{=} \left(g \circ \phi^{-1}\right)_{\phi(q)}. \tag{6.15}$$

In order to use the Inverse Function Theorem for maps on manifolds, the notion of regularity must be extended to abstract maps. This we now do.

Definition 6.4: *Let $f : W \subset \mathcal{N} \to \Re^m$ be a C^r map, $r \geq 1$, $\mathcal{N}$ a differentiable n-manifold with $n \leq m$. f is said to be regular at $q \in W$ if there exists a coordinate neighborhood (U, ϕ) of q such that $\hat{f} = f \circ \phi^{-1}$ is regular at q in the sense of Definition 6.1. f is said to be regular if it is regular at every $q \in W$.*

A regular C^r map is called a C^r *immersion.* If $r = \infty$, then the map is called simply an *immersion.* It is important to note that a C^r immersion need not be 1-1. However, if a C^r immersion is 1-1, then it is called a C^r *embedding.* The next theorem shows that immersions are locally embeddings.

Theorem 6.5: *Let $f : \mathcal{N} \to \Re^m$ be a C^r immersion. Each $p \in \mathcal{N}$ has a neighborhood U such that f restricted to U is a C^r embedding. That is, f is locally $1-1$.*

Proof: For a proof, see Boothby [14], Theorem 4.12. ∎

The notion of a diffeomorphism can be extended readily to abstract maps on manifolds as follows. Let $f : W \to \Re^m$, $W \subset \mathcal{N}$, and let W' be the image of f; that is $W' = f(W)$. f is said to be a C^r-*diffeomorphism* if f is C^r, and $f^{-1} : W' \to W$ exists and is C^r. The next theorem shows the connection between embeddings and diffeomorphisms.

Theorem 6.6: *If $f : W \to \Re^m$, $W \subset \mathcal{N}$ is a C^r embedding, then f is a C^r-diffeomorphism.*

Proof: For a proof, see Boothby [14], Remark 4.2. ∎

Remark 6.6: *Theorems 6.5 and 6.6 together imply that immersions are locally diffeomorphisms.*

The next theorem shows that the image of a manifold under an embedding is itself a manifold, and is called an *embedded manifold.*

Theorem 6.7: *Let $\mathcal{N}$ be a differentiable manifold, and let $f : \mathcal{N} \to \Re^m$ be a C^r embedding, $r = 1, 2, ..., \infty$, where $\dim(\mathcal{N}) \leq m$. The image of $\mathcal{N}$ under f, $f(\mathcal{N})$, is a C^r manifold.*

Proof: For a proof, see Boothby [14], Chapter 4. ∎

We will say that a manifold $\mathcal{N}$ is *compact* if there exists an embedding $f : \mathcal{N} \to \Re^m$, for sufficiently large m, such that $f(\mathcal{N})$ is compact when considered as a subset of $\Re^m$. Once again, this is a narrow definition of compactness, but it suffices for our purposes.

Remark 6.7: *Theorems 6.5 and 6.7 together imply that the image of a C^r immersion is locally an embedded manifold. That is, if f is a C^r immersion defined on $\mathcal{N}$, then for every $q \in \mathcal{N}$ there exists a neighborhood U of q such that $f(U)$ is an embedded C^r manifold. The image set $f(\mathcal{N})$ is called an immersed manifold.*

6.5. Abstract Splines

As mentioned in Section 6.1, the notion of an abstract spline is useful when developing a coordinate free measure of continuity. Intuitively, an abstract spline is a coordinate free version of the usual notion of a spline as a piecewise map. The general idea is to start with a manifold, "slice" it into smaller sub-domains, then on each sub-domain define a map into Euclidean space.

In order to define an abstract spline on a manifold domain, we first introduce the notion of a tesselation, used to slice the domain manifold into closed sub-domains.

Definition 6.5: *Let $\mathcal{P}$ be a compact C^∞ manifold with boundary, let* $\dim \mathcal{P} = p$, *and let $\partial\mathcal{P}$ denote the boundary of $\mathcal{P}$. A tesselation of $\mathcal{P}$ (if it exists) is a finite collection Φ of subsets of $\mathcal{P}$ such that:*

i) Elements of Φ (called sub-domains) cover $\mathcal{P}$.

ii) Elements of Φ are homeomorphic to $\bar{\mathbf{B}}_1^p(0)$, the closure of a ball in $\Re^p$ of radius 1 about the origin.

iii) If $q \in \mathcal{P}$ is an interior point of some $D \in \Phi$, then no other $D' \in \Phi$ contains q.

iv) For every $D \in \Phi$ there is a coordinate neighborhood (U, ψ) of $\mathcal{P}$ such that $D - (D \cap \partial\mathcal{P}) \subset U$. U is called an extension of D.

v) If $D \cap D' \neq \emptyset$, then $D \cap D'$ is a differentiable manifold of dimension less than p.

A tesselated manifold is a tuple $(\mathcal{P}, \Phi)$, $\mathcal{P}$ a manifold, Φ a tesselation of $\mathcal{P}$.

Remark 6.8: *Since Φ is a finite collection, it can be ordered: $\Phi = \{D_1, ..., D_t\}$, $1 < t < \infty$. Also, since the elements of Φ cover $\mathcal{P}$, the coordinate neighborhoods*

$\{(U_i, \psi_i)\}$, U_i *an extension of* D_i, *form a basis for* $\mathrm{Int}(\mathcal{P})$.

Example 6.3: *Let* $\mathcal{P}$ *be the closed subset of* $\Re^2$ $[0,2] \times [0,2]$, *and let* $\Phi = \{D_1, D_2, D_3, D_4\}$ *be a rectangular tesselation of* $\mathcal{P}$. *In particular, let* $D_1 = [0,1] \times [0,1]$, $D_2 = [0,1] \times [1,2]$, $D_3 = [1,2] \times [1,2]$, *and* $D_4 = [1,2] \times [0,1]$ *(see Figure 6.6). Note that* D_1 *and* D_2 *intersect in a 1-manifold (a straight line), and* D_1 *and* D_3 *intersect in a 0-manifold (a single point). It is easily verified that* Φ *forms a tesselation of* $\mathcal{P}$; *hence,* $(\mathcal{P}, \Phi)$ *is a tesselated 2-manifold.* •

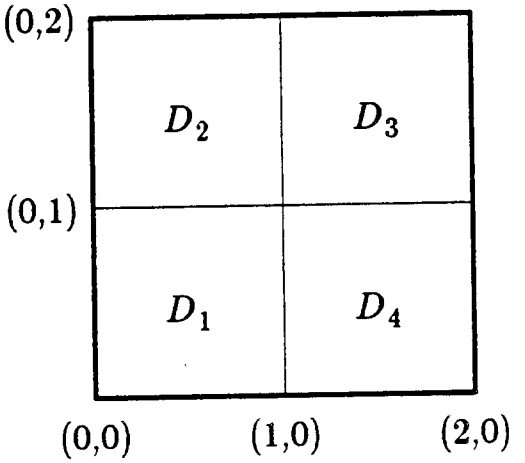

Figure 6.6. *A rectangular tesselation of the manifold* $[0,2] \times [0,2]$.

Definition 6.6: *Let* $(\mathcal{P}, \Phi)$ *be a tesselated manifold, with* $\Phi = \{D_1, ..., D_t\}$. *Associate with each* $D_i \in \Phi$ *a* C^∞ *map* $s_i : U_i \to \Re^m$, U_i *an extension of* D_i. *Define a piecewise map* $S : \mathrm{Int}(\mathcal{P}) \to \Re^m$ *as*

$$S(q) = s_i(q), \qquad q \in \mathrm{Int}(\mathcal{P}), \qquad i = \min\{j | q \in D_j\}. \tag{6.16}$$

S *is called an abstract spline on* $(\mathcal{P}, \Phi)$; S *is said to be given by the collection* $\{s_i, D_i\}$. S *is said to be regular if each of the* s_i *is regular.*

It is easy to show that S is indeed a map on $\mathrm{Int}(\mathcal{P})$ (see Figure 6.7). However, until we impose some restrictions on the maps s_i, S need not be differentiable or even continuous. To investigate the form of these restrictions, let S be a spline on $(\mathcal{P}, \Phi)$ given by $\{s_i, D_i\}$, and let q be a point on $\mathrm{Int}(\mathcal{P})$ that is on the boundary between D_1 and D_2, but not on the boundary of any other sub-domain. We ask the question: If S is C^r at q, what does that imply about the relationship between

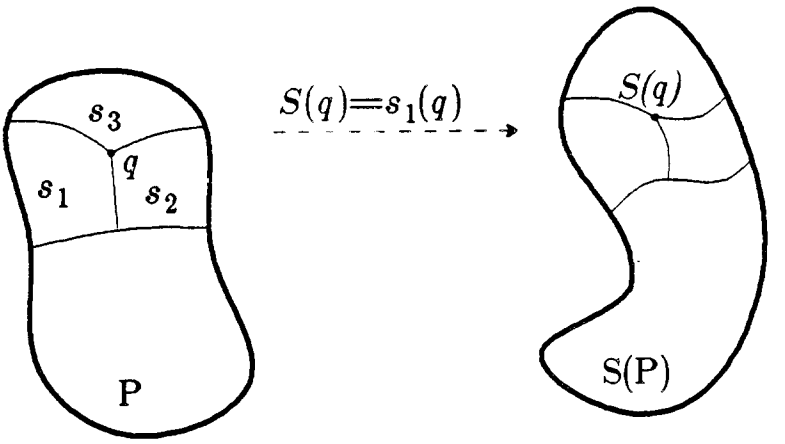

Figure 6.7. *If $q \in \text{Int}(P)$ is on the interior of a sub-domain D_i, then the value of the spline S is $s_i(q)$. If q is on the boundary between one or more sub-domains, the map on the sub-domain of least index is used to define the value of $S(q)$. In the figure above, q is on the boundary between D_1, D_2, and D_3, so $S(q) = s_1(q)$.*

s_1 and s_2? The fact that S is C^r at q implies that there exists a coordinate neighborhood (W, λ) of q such that the representation of S relative to (W, λ) is C^r at $\lambda(q)$. If $\widehat{S}$ denotes this representation, then

$$\widehat{S}(x) = S \circ \lambda^{-1}(x), \qquad x \in \lambda(W) \tag{6.17}$$

where $\widehat{S}$ is C^r at $\lambda(q)$ (see Figure 6.8).

Since S is defined as a piecewise map, so is $\widehat{S}$. Specifically,

$$\widehat{S}(x) = \begin{cases} s_1 \circ \lambda^{-1}(x) & x \in \lambda(D_1) \\ s_2 \circ \lambda^{-1}(x) & \text{otherwise.} \end{cases} \tag{6.18}$$

By definition, $\widehat{S}$ is C^r at $\lambda(q)$ if and only if all partial derivatives up to order r exist and are continuous at $\lambda(q)$. Equivalently, $\widehat{S}$ is C^r at $\lambda(q)$ if and only if

$$\left(s_1 \circ \lambda^{-1}\right)_{\lambda(q)} \overset{C^r}{=} \left(s_2 \circ \lambda^{-1}\right)_{\lambda(q)}. \tag{6.19}$$

In general, any number of domains can contain q. The next claim handles the general case. It can be proved by using an analysis identical to the one used above.

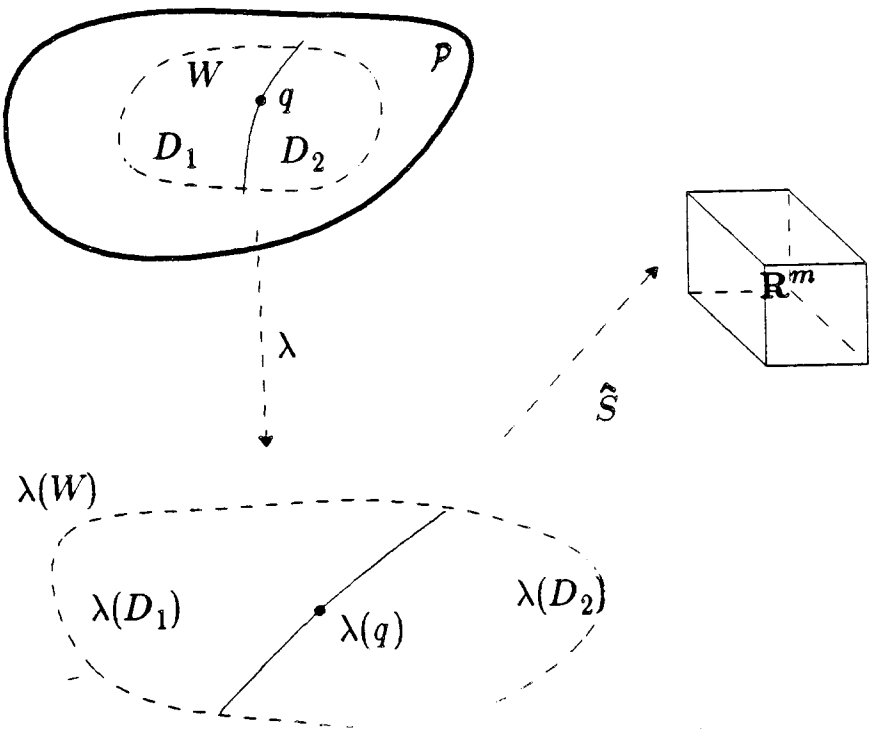

Figure 6.8. *The representation of a spline S relative to a coordinate neighborhood (W, λ) for a point $q \in \mathrm{Int}(P)$ on the boundary between D_1 and D_2.*

Claim 6.2: *Let S be an abstract spline on (P, Φ) given by $\{s_i, D_i\}$, and let (W, λ) be a coordinate neighborhood of $q \in \mathrm{Int}(P)$. Let $\widehat{S}(x)$, $x \in W$, denote the representation of S relative to (W, λ):*

$$\widehat{S}(x) = s_i \circ \lambda^{-1}(x), \qquad i = \min\{j | x \in \lambda(D_j)\},$$

and let $D_{i_1}, ..., D_{i_k}$ be all domains containing q. $\widehat{S}$ is C^r at $\lambda(q)$ if and only if

$$\left(s_{i_1} \circ \lambda^{-1}\right)_{\lambda(q)} \overset{C^r}{=} \left(s_{i_2} \circ \lambda^{-1}\right)_{\lambda(q)} \overset{C^r}{=} \cdots \overset{C^r}{=} \left(s_{i_k} \circ \lambda^{-1}\right)_{\lambda(q)} \qquad (6.20)$$

Remark 6.9: *Expression (6.20) holds if and only if it holds on a pairwise basis.*

Abstract splines are convenient theoretic tools, but in practice we must deal with representations. The representation $\widehat{S}$ in Claim 6.2 is an example of a *local representation* of S. The representation is local in the sense that it is only valid for the neighborhood W of q. A *global representation* of S can be considered to be a collection of local representations. Of course, there are many possible global

representations, but there is a class of global representations that is particularly useful in practice. The idea is to use global representations that consist of a collection of representations of the maps s_i. We now formalize this idea.

Let $\hat{D}$ be a subset of $\Re^p$ that is homeomorphic to $\bar{\mathbf{B}}_1^p(0)$, let $\hat{U}$ be an open subset of $\Re^p$ containing $\hat{D}$, let $x = (x_1, ..., x_p)$ be a coordinate system on $\hat{U}$, and let $\hat{s}(x)$ be a C^∞ function from $\hat{U}$ into $\Re^m$. $\hat{s}$, when restricted to $\hat{D}$, is called a *parametrization*, and is said to be *parametrized on* $\hat{D}$. The tuple $(\hat{s}, \hat{D})$ is used to denote a parametrization $\hat{s}$ parametrized on $\hat{D}$.

Definition 6.7: *Let S be an spline on $(\mathcal{P}, \Phi)$ given by $\{s_i, D_i\}$. Let (U_i, ψ_i) be a coordinate neighborhood such that s_i is defined on U_i, for U_i an extension of D_i. A parametrization $(\hat{s}_i, \hat{D}_i)$ is said to be a parametrization for s_i relative to (U_i, ψ_i) if*

$$\hat{s}_i(x) = s_i \circ \psi_i^{-1}(x), \qquad x \in \psi_i(U_i) \tag{6.21}$$

and $\hat{D}_i = \psi(D_i)$. That is, $\hat{s}_i$ is a representation of s_i relative to (U_i, ψ_i). Conversely, s_i is said to be a lifting of $(\hat{s}_i, \hat{D}_i)$ relative to (U_i, ψ_i). The coordinate neighborhood (U_i, ψ_i) is called a connecting neighborhood (intuitively, it "connects" s_i and $(\hat{s}_i, \hat{D}_i)$).

Let S be as above, and let $\{(U_i, \psi_i)\}$ be a collection of coordinate neighborhoods such that s_i is defined on U_i, U_i an extension of D_i. A parametric representation of S relative to $\{(U_i, \psi_i)\}$ is a collection of parametrizations $\{(\hat{s}_i, \hat{D}_i)\}$ such that for each i, s_i is a lifting of $(\hat{s}_i, \hat{D}_i)$ relative to (U_i, ψ_i). The set of connecting neighborhoods $\{(U_i, \psi_i)\}$ relating S and $\{(\hat{s}_i, \hat{D}_i)\}$ is called a connecting basis.

A collection $\{(\hat{s}_i, \hat{D}_i)\}$ is said to be a parametric representation on $(\mathcal{P}, \Phi)$ if the collection is a parametric representation of some abstract spline S on $(\mathcal{P}, \Phi)$. Conversely, S is said to be a lifting of $\{(\hat{s}_i, \hat{D}_i)\}$.

6.6. Parametric Splines

As mentioned in Section 6.1, a parametric representation is "lifted" onto the manifold to define a coordinate free abstract spline. However, not just any collection of parametrizations forms a parametric representation. The parametric domains, when lifted onto the manifold, must fit together (without overlapping) to cover the manifold. The next example should help to clarify some of these ideas.

Example 6.4: *We examine the relationship between a parametric representation and an associated abstract spline for the surface shown in Figure 6.9. We begin*

by discussing the construction in the old language of Chapter 2, and then show how the construction relates to the new language of abstract splines on manifolds.

The surface shown in Figure 6.9 is represented by the four parametrizations $\hat{s}_1, \hat{s}_2, \hat{s}_3$, *and* $\hat{s}_4$. $\hat{s}_1$ *is parametrized on* $[0,1] \times [0,2]$, $\hat{s}_2$ *on* $[0,2] \times [0,2]$, *and* $\hat{s}_3$ *and* $\hat{s}_4$ *on* $[0,1] \times [0,1]$.

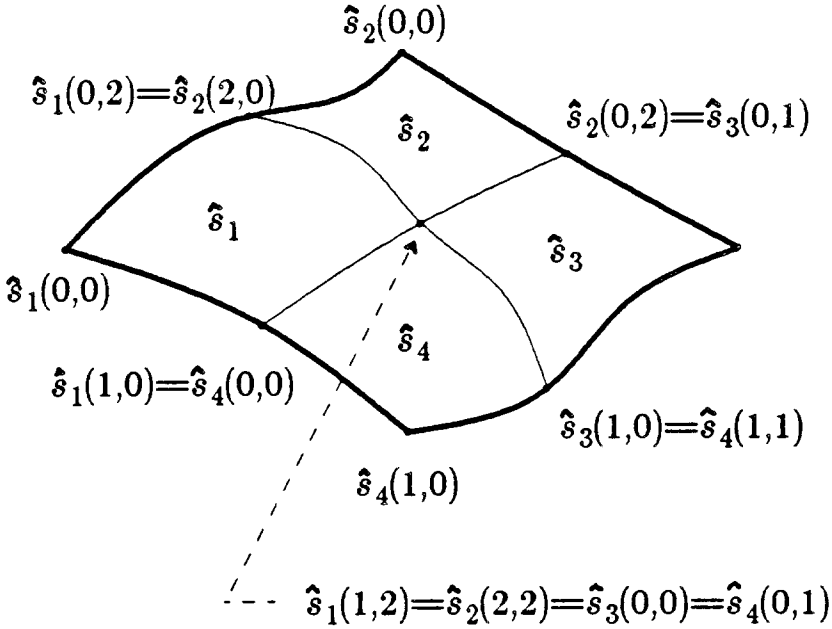

Figure 6.9. *A spline surface consisting of four patches generated by* $\hat{s}_1$, $\hat{s}_2$, $\hat{s}_3$ *and* $\hat{s}_4$.

As demonstrated in Section 2.5, if the surface is to be continuous (C^0), *a set of correspondence maps must be established that relate parametric boundary points. We choose the simplest correspondence maps consistent with the situation shown in Figure 6.9. The correspondence map between* $\hat{s}_1$ *and* $\hat{s}_4$ *is is partially established by Figure 6.9 in that the point* $\hat{s}_1(1,0)$ *is associated with the point* $\hat{s}_4(0,0)$, *and* $\hat{s}_1(1,2)$ *is associated with* $\hat{s}_4(0,1)$. *Equivalently,* $(1,0)$ *in* $\hat{s}_1$*'s domain is associated with* $(0,0)$ *in* $\hat{s}_4$*'s domain, and* $(1,2)$ *in* $\hat{s}_1$*'s domain is associated with* $(0,1)$ *in* $\hat{s}_4$*'s domain. Let* (u_1, v_1) *be the coordinate system on* $\hat{s}_1$*'s domain, and let* (u_4, v_4) *be the coordinate system on* $\hat{s}_4$*'s domain. The simplest correspondence along the* $\hat{s}_1, \hat{s}_4$ *boundary consistent with Figure 6.9 is therefore*

$$\begin{aligned} u_4 &= u_1 - 1 \\ v_4 &= \frac{1}{2} v_1 . \end{aligned} \tag{6.22}$$

Similarly, the simplest correspondence along the $\hat{s}_2, \hat{s}_3$ boundary is

$$\begin{aligned} u_3 &= v_2 - 2 \\ v_3 &= \frac{1}{2}(2 - u_2) \end{aligned} \tag{6.23}$$

where (u_2, v_2) and (u_3, v_3) are coordinates systems on the domains of $\hat{s}_2$ and $\hat{s}_3$, respectively (note that even the simplest correspondence of equation (6.23) mixes the u's and v's of the coordinate systems). The correspondences for the remaining two boundaries can be determined in a similar fashion.

The C^0 condition along the $\hat{s}_1, \hat{s}_4$ boundary can now be stated as

$$\begin{aligned} \hat{s}_1(1, v_1) &= \hat{s}_4(u_4 = 1 - 1, v_4 = \frac{1}{2}v_1) \\ &= \hat{s}_4(0, \frac{1}{2}v_1) \end{aligned} \tag{6.24}$$

for all $v_1 \in [0, 2]$. Similar expressions for the other boundaries can be obtained analogously, thus completing the construction of a C^0 surface in the old language.

To demonstrate how the parametric representation given above can be lifted onto a manifold, we choose a tesselated manifold that admits an abstract spline having $\{\hat{s}_1, \hat{s}_2, \hat{s}_3, \hat{s}_4\}$ as a parametric representation. In particular, we choose the tesselated manifold of Example 6.3, lifting $\hat{s}_1$ onto D_1, $\hat{s}_2$ onto D_2, and so on. We must now choose a connecting basis; i.e., the coordinate neighborhoods (U_1, ψ_1), (U_2, ψ_2), (U_3, ψ_3), and (U_4, ψ_4) relative to which the lifting will occur. For the extensions U_1, U_2, U_3, U_4, simply choose $U_1 = U_2 = U_3 = U_4 = (0, 2) \times (0, 2)$. $\hat{s}_1$ is parametrized on $[0, 1] \times [0, 2]$, so ψ_1 must be such that $\psi_1(D_1) = [0, 1] \times [0, 2]$. Similarly, $\psi_2(D_2) = [0, 2] \times [0, 2]$, $\psi_3(D_3) = [0, 1] \times [0, 1]$, and $\psi_4(D_4) = [0, 1] \times [0, 1]$ (see Figure 6.10).

The fact that $(1, 0)$ in $\hat{s}_1$'s domain is associated with $(0, 0)$ in $\hat{s}_4$'s domain means that $(1, 0)$ and $(0, 0)$ are images of the same point on the manifold. In other words, the same point on the manifold, and hence the same point on the spline surface, is given a different coordinate assignment relative to different coordinate systems, or equivalently, relative to different parametrizations. The conversion from $\hat{s}_1$'s coordinate to $\hat{s}_4$'s coordinate for a point on the $\hat{s}_1, \hat{s}_4$ boundary is given by equation (6.22). In the new language, this correspondence map describes the transition function $\psi_4 \circ \psi_1^{-1}$ restricted to the $u_1 = 1$ edge of $\hat{s}_1$'s domain. Thus, the C^0 condition of equation (6.24) can be rewritten in the new language as

$$(\hat{s}_1)_{\psi_1(q)} = \left(\hat{s}_4 \circ \psi_4 \circ \psi_1^{-1}\right)_{\psi_1(q)}, \qquad \forall q \in D_1 \cap D_4. \tag{6.25}$$

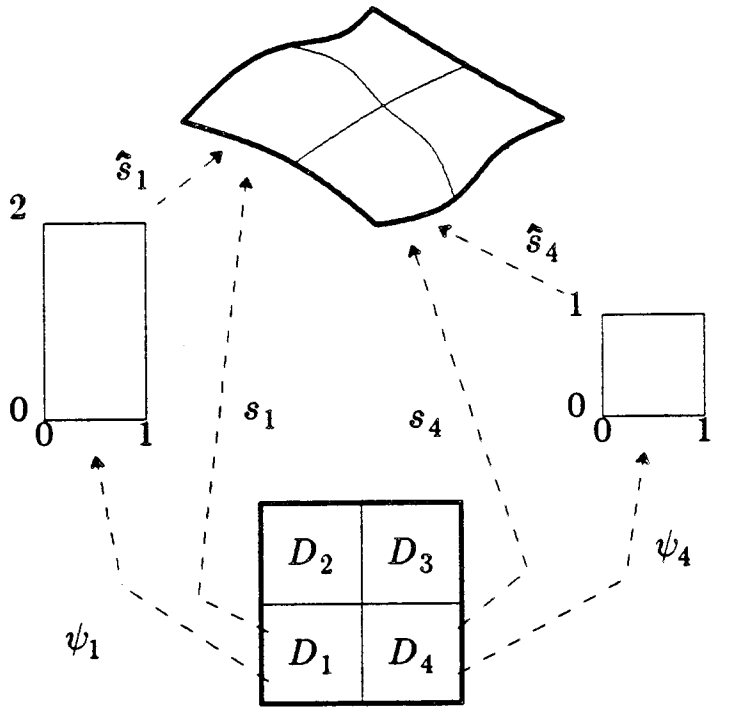

Figure 6.10. *The relationship between the manifold, the parametrizations, and the abstract spline.*

Equation (6.25) can be written in terms of abstract maps $s_1 = \hat{s}_1 \circ \psi_1$ *and* $s_4 = \hat{s}_4 \circ \psi_4$ *as*

$$\left(s_1 \circ \psi_1^{-1}\right)_{\psi_1(q)} = \left(s_4 \circ \psi_1^{-1}\right)_{\psi_1(q)}, \qquad \forall q \in D_1 \cap D_4. \tag{6.26}$$

Equation (6.26) is therefore the coordinate free form of continuity between $\hat{s}_1$ *and* $\hat{s}_4$. *The parametrizations* $\hat{s}_2$ *and* $\hat{s}_3$ *can be lifted in a similar manner.* •

Example 6.4 shows that when building splines in practice, one begins with a collection of parametrizations, together with a set of correspondence maps that ensure C^0 continuity. Our definition of a *parametric spline*, that is, a spline that is constructed in practice, should contain this information. There are many possible definitions that capture this information. The most obvious definition would be to treat a parametric spline as a collection of parametrizations together with a set of correspondence maps. Although conceptually simple, this definition does not seem to be particularly convenient for a theoretical study. The definition of a parametric spline we will ultimately use is rather esoteric, so we begin with some motivational observations.

Notice that the abstract spline in Example 6.4 is not uniquely defined since the charts with respect to which the lifting occurred were not unique; any other connecting basis consistent with the correspondence maps could have been used. Intuitively this means that a given parametric representation represents many abstract splines consistent with a fixed set of correspondence maps. To investigate this further, let S be a C^0 abstract spline on $(\mathcal{P}, \Phi)$, let $\{(\hat{s}_i, \hat{D}_i)\}$ be a parametric representation relative to the connecting basis $\{(U_i, \psi_i)\}$, let q be a point on the manifold common to the sub-domains D_i and D_j, and let (U_i, ψ_i), (U_j, ψ_j) be the connecting neighborhoods that are associated with D_i and D_j. Thus, $\psi_i(D_i) = \hat{D}_i$ and $\psi_j(D_j) = \hat{D}_j$. As shown in Example 6.4, the point $\psi_i(q)$ in $\hat{D}_i$ is associated with the point $\psi_j(q)$ in $\hat{D}_j$. In this way, a connecting basis uniquely determines a set of correspondence maps. However, the converse is not true — a set of correspondence maps does not uniquely determine a connecting basis. In particular, let ψ_i' be a chart on U_i such that

$$\begin{aligned} &1)\ \psi_i'(D_i) = \psi_i(D_i) \\ &2)\ \psi_i'(q) = \psi_i(q), \text{ for all } q \in \partial D_i. \end{aligned} \tag{6.27}$$

That is, ψ_i and ψ_i' behave identically on the boundary of D_i. Two charts satisfying properties 1 and 2 of equation (6.27) are said to be *similar* (actually, 2 implies 1 since D_i is homeomorphic to $\mathbf{B}_1^p(0)$). The interesting thing about similar charts is that they preserve correspondence maps. That is, $\psi_i(q)$ corresponds to $\psi_j(q)$ if and only if $\psi_i'(q)$ corresponds to $\psi_j(q)$. In addition, properties 1 and 2 of equation (6.27) imply that $\hat{s}_i$ can be lifted relative to (U_i, ψ_i') to describe the abstract map $s_i' = \hat{s}_i \circ \psi_i'$. The abstract map s_i' is therefore related to the abstract map $s_i = \hat{s}_i \circ \psi_i$ (defined by the connecting basis), according to

$$s_i' = s_i \circ \psi_i^{-1} \circ \psi_i'. \tag{6.28}$$

Equation (6.28) shows that $\hat{s}_i$ is a representation of s_i relative to ψ_i, as well as being a representation of s_i' relative to ψ_i'. Two abstract maps are said to be *similar* if they are related by similar charts as indicated by equation (6.28). Two splines S and S', given by $\{s_i, D_i\}$ and $\{s_i', D_i\}$ are said to be *similar* if for each i, s_i is similar to s_i'. Note that if S and S' are similar, then they have a common parametric representation, and if $\{(\hat{s}_i, \hat{D}_i)\}$ is a parametric representation, it is relatively easy to show that "similarity" is an equivalence relation on the set of abstract splines represented by $\{(\hat{s}_i, \hat{D}_i)\}$.

The important point of the above discussion is that a parametric representation and a set of correspondence maps determines an abstract spline only up to similarity. This observation is the basis of our definition of parametric splines.

Definition 6.8: *Let $\{(\hat{s}_i, \hat{D}_i)\}$ be a parametric representation on $(\mathcal{P}, \Phi)$, let S be a C^0 lifting of $\{(\hat{s}_i, \hat{D}_i)\}$, and let $\mathcal{C}$ be the equivalence class of abstract splines similar to S. The tuple $\widetilde{S} = (\{(\hat{s}_i, \hat{D}_i)\}, \mathcal{C})$ is called a parametric spline on $(\mathcal{P}, \Phi)$; $\widetilde{S}$ is said to represent S' if $S' \in \mathcal{C}$; $\widetilde{S}$ is said to be regular if each of the maps $\hat{s}_i$ is regular on $\hat{D}_i$. We will often be lax and use the term "spline" to be synonymous with "regular parametric spline".*

Remark 6.10: *With each $S \in \mathcal{C}$, there is an associated unique connecting basis relating S and $\{(\hat{s}_i, \hat{D}_i)\}$. Thus, an equivalent definition of a parametric spline is a tuple $(\{(\hat{s}_i, \hat{D}_i)\}, E)$, E an equivalence class of similar connecting bases. Similar connecting bases are defined in the obvious way: two connecting bases $\{(U_i, \psi_i)\}$ and $\{(U_i, \psi_i')\}$ are called similar if for each i, ψ_i is similar to ψ_i'. In fact, given any two of $\mathcal{C}$, E and $\{(\hat{s}_i, \hat{D}_i)\}$, the third is uniquely determined. Unfortunately, none of the $\binom{3}{2}$ possible definitions seems to be superior in all cases.*

Notation: *Let $\widetilde{S} = (\{(\hat{s}_i, \hat{D}_i)\}, \mathcal{C})$ be a parametric spline on $(\mathcal{P}, \Phi)$. We use the notation $\widetilde{S} : \mathrm{Int}(\mathcal{P}) \to \Re^m$ to mean that the abstract splines represented by $\widetilde{S}$ are maps from $\mathrm{Int}(\mathcal{P})$ into $\Re^m$. In a similar spirit, we take $\widetilde{S}(\mathrm{Int}(\mathcal{P}))$ to be the set $S(\mathrm{Int}(\mathcal{P}))$, $S \in \mathcal{C}$.*

As a follow up to Remark 2.2, it is important to realize that the assumption of C^0 continuity has been woven into the definition of a parametric spline by requiring all liftings to be C^0 abstract splines. The C^0 assumption is very convenient when working on the manifold. As a specific example of the difficulties encountered when the C^0 assumption is not made, consider the situation depicted in Figure 6.11. The parametric domains $\hat{D}_1, \hat{D}_2$, and $\hat{D}_3$ are to be lifted onto the manifold $[0, 3]$, tesselated into the sub-domains $[0, 1]$, $[1, 2]$, and $[2, 3]$. Since all closed intervals are diffeomorphic to one another, it is possible to lift any of the $\hat{D}_i$ onto any of the sub-domains. In fact, each of the liftings can be performed in either of two possible ways. Without the C^0 assumption, it is unclear which of the many liftings will result in a C^0 spline; three possibilities are shown in Figure 6.11. In Figure 6.11(a), $\hat{D}_1$ and $\hat{D}_2$ are "adjacent" on the manifold, with the point $1 \in \hat{D}_1$ corresponding to $0 \in \hat{D}_2$. $\hat{D}_1$ and $\hat{D}_2$ are again adjacent in Figure 6.11(b), except that $0 \in \hat{D}_1$ corresponds to $0 \in \hat{D}_2$. In Figure 6.11(c), $\hat{D}_1$ and $\hat{D}_2$ aren't even adjacent. Thus, without the C^0 assumption, even the notion of adjacency is lost.

However, when one constructs splines in practice, there is a very definite idea of which parametrizations are to be treated as adjacent — the C^0 assumption provides this structure while still allowing cross boundary freedom.

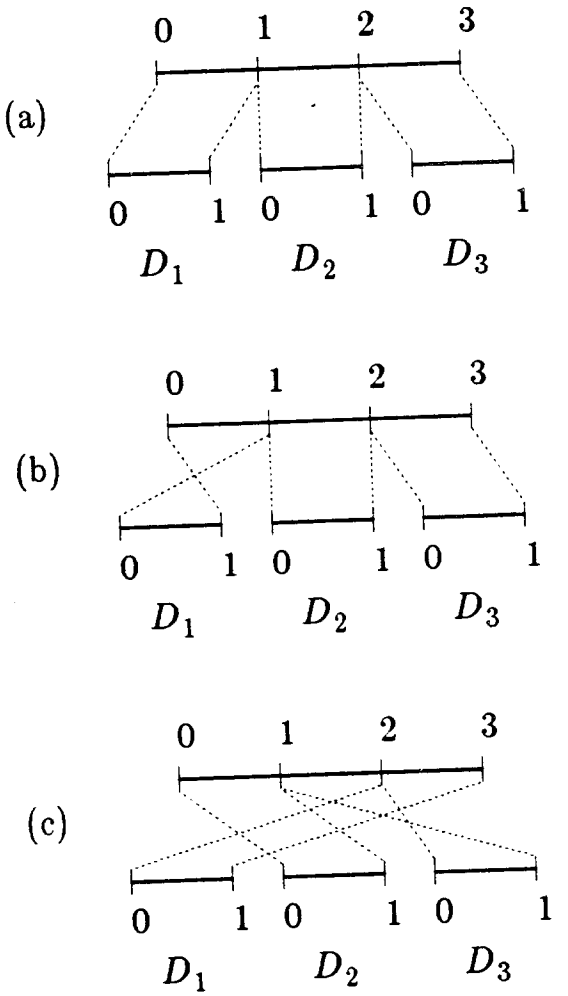

Figure 6.11. *Three possible liftings of* $[0,1]$, $[0,1]$, $[0,1]$ *onto the tesselated manifold* $([0,3],\{[0,1],[1,2],[2,3]\})$.

6.7. Weak Geometric Continuity, WG^r Splines

As mentioned previously, an equivalence class of abstract splines is used to stitch the domains of a parametric spline together to achieve C^0 continuity. The goal is now to determine parametrization independent constraints on the $\hat{s}_i$ that

are necessary and sufficient for higher order differentiability of an abstract spline S represented by the parametric spline. This can be achieved by:

1) Lifting the representations $\hat{s}_i$ onto the manifold $\mathcal{P}$ to obtain abstract maps s_i.
2) Using the usual definition (Definition 6.3) of continuity for a map on a manifold. Since this is a measure of continuity that holds independent of the local coordinate system, it is independent of the representation of the s_i.
3) Restate the constraints on the manifold in terms of the parametrizations $\hat{s}_i$.

The underlying mathematical structure of a parametric spline provides the key to step 1, and step 3 is simply the inverse of step 1. The key to step 2 is provided by Claim 6.2, so we must simply put the pieces together, a task we now undertake.

It was shown above that equation (6.19) is a necessary and sufficient condition on the abstract maps s_1 and s_2 if S is to be C^r. From Claim 6.1, we see that equation (6.19) is a coordinate independent statement — it is therefore the key to a parametrization independent measure of continuity for splines. Formally:

Definition 6.9: *Let $\widetilde{S} = (\{(\hat{s}_i, \widehat{D}_i)\}, \mathcal{C})$ be a regular parametric spline on $(\mathcal{P}, \Phi)$, and let S be a lifting of $\widetilde{S}$ relative to $\{(U_i, \psi_i)\}$. The parametrizations $\hat{s}_i$ and $\hat{s}_j$ are said to abut at q if $q \in D_i \cap D_j$. $\hat{s}_i$ and $\hat{s}_j$ are said to abut on V if $V \subset D_i \cap D_j$.*

Let $\widetilde{S}$ be as above, let $\hat{s}_i$ and $\hat{s}_j$ abut on V, and let $s_i = \hat{s}_i \circ \psi_i$, $s_j = \hat{s}_j \circ \psi_j$, then:

- *$\hat{s}_i$ and $\hat{s}_j$ are said to meet with WG^r continuity at $q \in V$ with respect to S if*

$$(s_i \circ \psi_i^{-1})_{\psi_i(q)} \overset{C^r}{=} (s_j \circ \psi_i^{-1})_{\psi_i(q)} \tag{6.29}$$

- *$\hat{s}_i$ and $\hat{s}_j$ are said to meet with WG^r continuity at $q \in V$ if there exists a lifting S with respect to which they meet WG^r at q.*
- *$\widetilde{S}$ is said to be a WG^r spline at q with respect to S if all parametrizations abutting at q meet with WG^r continuity at q with respect to S.*
- *$\widetilde{S}$ is said to be a WG^r spline on $(\mathcal{P}, \Phi)$ if there exists a lifting S with respect to which $\widetilde{S}$ is WG^r at all $q \in \mathrm{Int}(\mathcal{P})$.*

We call this "weak" geometric continuity because no stipulations have yet been made concerning orientation. The orientation preserving conditions will be developed in Section 6.8.

Remark 6.11: *In Remark 2.1, it was pointed out that there is a possible alternate definition for two parametrizations meeting with geometric continuity on a set. Within the framework of manifolds, and using the definition above for $\hat{s}_i$ and $\hat{s}_j$ meeting with WG^r continuity at a point q, an analogous definition for weak geometric continuity may be stated as: $\hat{s}_i$ and $\hat{s}_j$ meet with WG^r continuity on $W \subset V$ if they meet with WG^r continuity at all $q \in W$. This definition of WG^r on a set allows the lifting used at $q_1 \in W$ to differ from the lifting used at $q_2 \in W$. It is conjectured that the two definitions are in fact equivalent, but a proof has not been constructed.*

The alternate definition has the advantage that continuity on a set can be guaranteed by individually establishing continuity at each point of the set, eliminating a consideration of the set as a whole. Thus, the alternate definition is better suited to proving Theorem 6.12 (the Beta constraint theorem), and the equivalence theorems given in Section 6.10. We have chosen to adopt Definition 6.9 because it expedites the proofs of the theorems detailing the smoothness of the images of WG^r splines (Theorems 6.8 and 6.9).

The next lemma establishes several minor (but useful) equivalent definitions of weak geometric continuity at a point. In particular, it shows that Definition 6.9 is reflexive in the sense that ψ_j could have been used instead of ψ_i in equation (6.29). In Section 6.10, additional equivalent definitions are proven.

Lemma 6.1: *Let $\widetilde{S} = (\{(\hat{s}_i, \widehat{D}_i)\}, \mathcal{C})$ be a parametric spline on $(\mathcal{P}, \Phi)$, and let S be a lifting of $\widetilde{S}$ relative to $\{(U_i, \psi_i)\}$. The following expressions are necessary and sufficient conditions for WG^r continuity of $\hat{s}_i$ and $\hat{s}_j$ with respect to S at q:*

$$\left(s_i \circ \psi_j^{-1}\right)_{\psi_j(q)} \overset{C^r}{=} \left(s_j \circ \psi_j^{-1}\right)_{\psi_j(q)} \tag{6.30}$$

$$\left(\hat{s}_i\right)_{\psi_i(q)} \overset{C^r}{=} \left(\hat{s}_j \circ \psi_j \circ \psi_i^{-1}\right)_{\psi_i(q)} \tag{6.31}$$

$$\left(\hat{s}_j\right)_{\psi_j(q)} \overset{C^r}{=} \left(\hat{s}_i \circ \psi_i \circ \psi_j^{-1}\right)_{\psi_j(q)} \tag{6.32}$$

where $s_i = \hat{s}_i \circ \psi_i$ and $s_j = \hat{s}_j \circ \psi_j$.

Proof: Equation (6.30) follows immediately from equation (6.29) and Claim 6.1.

Equation (6.31) follows from (6.29) by first using the fact that $\hat{s}_i = s_i \circ \psi_i^{-1}$. Specifically, (6.29) can be rewritten as

$$\left(\hat{s}_i\right)_{\psi_i(q)} \overset{C^r}{=} \left(s_j \circ \psi_i^{-1}\right)_{\psi_i(q)}. \tag{6.33}$$

Now, rewrite the right side of (6.33) by inserting the indentity in the form of $\psi_j^{-1} \circ \psi_j$:

$$(\hat{s}_i)_{\psi_i(q)} \overset{C^r}{=} \left(s_j \circ \psi_j^{-1} \circ \psi_j \circ \psi_i^{-1}\right)_{\psi_i(q)}. \tag{6.34}$$

Equation (6.34) can now be written as equation (6.31) by using $\hat{s}_j = s_j \circ \psi_j^{-1}$.

Equation (6.32) follows from (6.30) in the same way as (6.31) followed from (6.29). ∎

Let us digress for a moment to make a few comments concerning parametric continuity. Recall that parametric continuity, as introduced in Chapter 2, requires that parametric derivatives agree at common points between parametrizations. In the language of manifold theory, parametric continuity may be stated as follows.

Definition 6.10: *Let $\hat{s}_i$ and $\hat{s}_j$ be as in Definition 6.9. They meet with parametric continuity of order r (C^r) on V if they meet with WG^r on V and the Jacobian matrix of $\psi_j \circ \psi_i^{-1}$ is the identity matrix.*

Definition 6.10 shows that parametric continuity is very restrictive. It essentially states that $\hat{s}_i$ and $\hat{s}_j$ must share the same parameter space, or equivalently, the same local coordinate system. Thus, parametric continuity is a concept that is clearly not coordinate independent.

Remark 6.12: *The important point is this: In the old language, parametric continuity seemed natural, and geometric continuity was developed as a rather subtle extension. However, in the new language of manifolds we start with a coordinate free framework. In that framework, geometric continuity is natural, and parametric continuity is a subtle special case.*

In this section, the notion of a WG^r spline has been introduced as the type of spline that results when parametrizations are stitched together with weak geometric continuity. We now wish to characterize the behavior of these splines by examining the smoothness properties of their images. After all, it is the image of the spline that is relevant in CAGD. The next lemma and theorem show that WG^r splines and C^r immersions (introduced in Section 6.4.2) are, in a sense to be defined, really the same thing.

Lemma 6.2: *Let $\widetilde{S} = (\{(\hat{s}_i, \hat{D}_i)\}, \mathcal{C})$ be a spline on $(\mathcal{P}, \Phi)$. $\widetilde{S}$ represents a regular C^r map at $q \in \mathrm{Int}(\mathcal{P})$ if and only if $\widetilde{S}$ is a WG^r spline at q.*

Proof: Let $D_{i_1}, ..., D_{i_k}$ be all sub-domains of Φ containing q. We begin by assuming that $\widetilde{S}$ represents a regular C^r map at q, meaning that there exists an abstract spline $S \in \mathcal{C}$ that is a regular C^r map at q. Let $S \in \mathcal{C}$ be such an abstract spline, and let $\{(U_i, \psi_i)\}$ be the connecting basis relative to which S is a lifting of $\{(\hat{s}_i, \widehat{D}_i)\}$. S is therefore an abstract map given by $\{s_i, D_i\}$. Since S is C^r at q, Definition 6.3 implies that there exists a coordinate neighborhood (W, λ) of q such that $\widehat{S}(x) = S \circ \lambda^{-1}(x)$, $x \in \lambda(W)$ is C^r at $\lambda(q)$. Due to the coordinate independent nature of differentiability, this must hold for *any* chart λ defined on W. From Claim 6.2 and Remark 6.9, it must be that

$$\left(s_i \circ \lambda^{-1}\right)_{\lambda(q)} \overset{C^r}{=} \left(s_j \circ \lambda^{-1}\right)_{\lambda(q)} \tag{6.35}$$

holds for every pair of indices i, j chosen from $\{i_1, ..., i_k\}$. Now, restrict attention to an open subset

$$V = W \cap U_{i_1} \cap U_{i_2} \cdots \cap U_{i_k}.$$

Since V is a subset of W, equation (6.35) must also hold for any chart defined on V. In fact, V has been chosen so that each of the charts $\psi_{i_1}, ..., \psi_{i_k}$ is defined on V. In particular, let $\lambda = \psi_i$, the chart associated with $\hat{s}_i$. Equation (6.35) relative to ψ_i becomes

$$\left(s_i \circ \psi_i^{-1}\right)_{\psi_i(q)} \overset{C^r}{=} \left(s_j \circ \psi_i^{-1}\right)_{\psi_i(q)} \tag{6.36}$$

which by Definition 6.9 means that $\hat{s}_i, \hat{s}_j$ meet with WG^r continuity at q. The pair i, j was chosen arbitrarily so it must hold for every pair. By Definition 6.9, $\widetilde{S}$ is therefore WG^r at q.

To prove the converse, assume that $\widetilde{S}$ is WG^r at q, implying that every pair $\hat{s}_i, \hat{s}_j$ abutting at q meets with WG^r continuity at q with respect to some $S \in \mathcal{C}$. Let S be such a lifting of $\widetilde{S}$ relative to $\{(U_i, \psi_i)\}$. We must show that there exists a coordinate neighborhood (W, λ) of q such that S relative to (W, λ) is regular and C^r at $\lambda(q)$.

Let $W = U_{i_1} \cap \cdots \cap U_{i_k}$. On W, all the charts $\psi_{i_1}, ..., \psi_{i_k}$ are defined; arbitrarily choose ψ_{i_1}. By assumption, every pair $s_{i'} = \hat{s}_{i'} \circ \psi_{i'}$, $s_{j'} = \hat{s}_{j'} \circ \psi_{j'}$ from $\{s_{i_1}, ..., s_{i_k}\}$ satisfies

$$\left(s_{i'} \circ \psi_{i_1}^{-1}\right)_{\psi_{i_1}(q)} \overset{C^r}{=} \left(s_{j'} \circ \psi_{i_1}^{-1}\right)_{\psi_{i_1}(q)}. \tag{6.37}$$

By Claim 6.2, $\widehat{S} = S \circ \psi_{i_1}^{-1}$ must be C^r at $\psi_{i_1}(q)$, which in turn implies that S is C^r at q. Regularity follows from regularity of the $\hat{s}$'s. ∎

Theorem 6.8: *Let $\widetilde{S}$ be as in Lemma 6.2. $\widetilde{S}$ represents C^r immersion on $(\mathcal{P}, \Phi)$ if and only if it is a WG^r spline on $(\mathcal{P}, \Phi)$.*

Proof: This follows immediately from the definition of a C^r immersion, Definition 6.9 for WG^r splines, and Lemma 6.2. ∎

Remark 6.7 and Theorem 6.8 together imply that the image of a WG^r spline is a C^r immersed manifold. Moreover, if the spline is 1-1 as well as being WG^r, then its image is an embedded C^r manifold. This provides further confidence that (weak) geometric continuity is a desirable measure of smoothness for the image of a spline. The next theorem and the remarks thereafter provide even more evidence of the smoothness of WG^r splines.

Theorem 6.9: *Let* $\widetilde{S} : \text{Int}(\mathcal{P}) \rightarrow \Re^{p+1}$ *be a parametric spline on* $(\mathcal{P}, \Phi)$, $\dim(\mathcal{P}) = p$, *and let* S *be a lifting of* $\widetilde{S}$. $\widetilde{S}$ *is a* WG^r *spline if and only if* $S(\text{Int}(\mathcal{P}))$ *is locally the graph of a* C^r *function. That is,* $\widetilde{S}$ *is* WG^r *if and only if for every* $q \in \text{Int}(\mathcal{P})$, *there exists a neighborhood* U *of* q, *a neighborhood* U' *of* $S(q)$, *and a coordinate system* $(x_1, ..., x_{p+1})$ *for* $\Re^{p+1}$, *such that*

$$S(U) = \{(x_1, ..., x_p, f(x_1, ..., x_p)) | (x_1, ..., x_p) \in U'\} \tag{6.38}$$

for some C^r *function* f.

Proof: This theorem is proved in most texts on manifold theory for the case of a C^r immersion. Since WG^r splines represent immersions, the same proof holds true here. For the complete proof for immersions, the reader is referred to Boothby [14]. ∎

The case $p = 1$ in Theorem 6.9 corresponds to WG^r spline curves in two-space. We therefore have as a corollary to Theorem 6.9: The image of a WG^r spline curve in two-space is locally the graph of a C^r function $f(x)$. Intuition developed in calculus leads us to believe that graphs of differentiable functions are smooth; hence, the image of a WG^r spline is smooth. Another corollary to Theorem 6.9 corresponds to $p = 2$, and may be stated as: The image of a WG^r spline surface in three-space is locally the graph of a C^r bivariate function $f(x, y)$. Once again, intuition from calculus suggests that the images of WG^r spline surfaces are also smooth.

6.8. G^r Splines

We now move from WG^r continuity to G^r continuity by maintaining orientation information. This is done by requiring the connecting bases to be oriented.

Definition 6.11: *Let $\widetilde{S} = (\{(\hat{s}_i, \widehat{D}_i)\}, \mathcal{C})$ be a spline on $(\mathcal{P}, \Phi)$, let $S \in \mathcal{C}$ be a lifting relative to $\{(U_i, \psi_i)\}$, and let $\hat{s}_i$ and $\hat{s}_j$ be two parametrizations abutting at $q \in \mathrm{Int}(\mathcal{P})$.*

- *$\hat{s}_i$ and $\hat{s}_j$ are said to meet with G^r continuity at q with respect to S if they meet with WG^r continuity at q with respect to S and (U_i, ψ_i) and (U_j, ψ_j) are coherently oriented.*
- *$\hat{s}_i$ and $\hat{s}_j$ are said to meet with G^r continuity at q if there exists a lifting $S \in \mathcal{C}$ with respect to which they meet with G^r continuity at q.*
- *$\widetilde{S}$ is said to be G^r at q with respect to S if every pair of parametrizations abutting at q meet with G^r continuity at q with respect to S.*
- *$\widetilde{S}$ is said to be a G^r spline (on $(\mathcal{P}, \Phi)$) if there exists a lifting $S \in \mathcal{C}$ with respect to which $\widetilde{S}$ is G^r at every $q \in \mathrm{Int}(\mathcal{P})$.*

The old language can be viewed as keeping the abstract spline fixed, while wandering through equivalent parametrizations to find some that meet with parametric continuity. On the other hand, the new view is to keep the parametrizations fixed, while wandering though the abstract splines. As will be shown in Theorem 6.10, the two views are equivalent when considering continuity between two parametrizations, but it is difficult to imagine how the old language would handle a entire collection of parametrizations. In the new language, a collection of parametrizations presents little difficulty since the manifold is used as a common platform on which the parametric domains are to be related.

In Chapter 2, reparametrization played an important role in the development of geometric continuity. In the new language, we say that two parametrizations $(\hat{s}, \widehat{D})$ and $(\tilde{s}, \widetilde{D})$ are GO-equivalent if there exists an orientation preserving diffeomorphism $d : \widetilde{D} \to \widehat{D}$ such that $\tilde{s} = \hat{s} \circ d$. More generally, we can say that they are *weakly GO-equivalent* (WGO-equivalent) if there is a diffeomorphism relating them, irrespective of its orientation properties. To develop some intuition for the formal mathematical structure of a parametric spline, the reader is encouraged to prove the following lemma as an "exercise".

Lemma 6.3: *Let $\widetilde{S} = (\{(\hat{s}_i, \widehat{D}_i)\}, \mathcal{C})$ be a regular parametric spline on $(\mathcal{P}, \Phi)$, and let $\{(\tilde{s}_i, \widetilde{D}_i)\}$ be a collection of parametrizations such that, for each i, $\hat{s}_i$ is WGO-equivalent to $\tilde{s}_i$. Show that $\widetilde{S}' = (\{(\tilde{s}_i, \widetilde{D}_i)\}, \mathcal{C})$ is a spline on $(\mathcal{P}, \Phi)$.*

Proof: Let $S \in \mathcal{C}$ be a lifting of $\{(\hat{s}_i, \widehat{D}_i)\}$ relative to $\{(U_i, \psi_i)\}$. Relative to this

connecting basis, $\hat{s}_i$ represents the abstract map

$$s_i = \hat{s}_i \circ \psi_i, \qquad D_i = \psi_i^{-1}(\hat{D}_i). \tag{6.39}$$

By assumption, $\{(\tilde{s}_i, \tilde{D}_i)\}$ is a collection of parametrizations such that for each i, $\tilde{s}_i = \hat{s}_i \circ d_i$, d_i a diffeomorphism such that $d_i(\tilde{D}) = \hat{D}$. Define a map ϕ_i by

$$\phi_i = d_i^{-1} \circ \psi_i. \tag{6.40}$$

By construction, ϕ_i is compatible with ψ_i, hence ϕ_i is a chart. From (6.39) and (6.40), we see that $s_i = \hat{s}_i \circ \psi_i = \tilde{s}_i \circ \phi_i$, implying that s_i is a lifting of $\tilde{s}_i$ relative to ϕ_i. This must be true for every i, so S is a lifting of $\{(\tilde{s}_i, \tilde{D}_i)\}$ relative to $\{(U_i, \phi_i)\}$. S is a representative of the equivalence class $\mathcal{C}$, so $\tilde{S}' = (\{(\tilde{s}_i, \tilde{D}_i)\}, \mathcal{C})$ is a spline on $(\mathcal{P}, \Phi)$. ∎

Note that $\tilde{S}$ and $\tilde{S}'$ of Lemma 6.3 are alike in the sense that they represent the same equivalence class of abstract splines, and therefore have the same image. However, they are different in the sense that they provide a different parametric representation for the equivalence class, implying that the correspondence maps differ.

The next theorem establishes the equivalence of the new and old views for continuity between two parametrizations.

Theorem 6.10: *Let $\tilde{S} = (\{(\hat{s}_i, \hat{D}_i)\}, \mathcal{C})$ be a regular parametric spline on $(\mathcal{P}, \Phi)$, and let $\hat{s}_i, \hat{s}_j$ be abut on $V \subset \mathcal{P}$. $\hat{s}_i$ and $\hat{s}_j$ meet with WG^r continuity on V if and only if there exist WGO-equivalent parametrizations $\tilde{s}_i$ and $\tilde{s}_j$, belonging to a spline $\tilde{S}' = (\{(\tilde{s}_i, \tilde{D}_i)\}, \mathcal{C})$ on $(\mathcal{P}, \Phi)$, that meet with C^r continuity on V. Moreover, $\hat{s}_i$ and $\hat{s}_j$ meet with G^r continuity on V if and only if there exist GO-equivalent parametrizations that meet with C^r continuity.*

Proof: We begin by assuming that $\hat{s}_i$ and $\hat{s}_j$ meet with WG^r continuity on V, implying that there exist connecting neighborhoods (U_i, ψ_i) and (U_j, ψ_j) such that

$$(\hat{s}_i)_{\psi_i(q)} \overset{C^r}{=} (\hat{s}_j \circ \psi_j \circ \psi_i^{-1})_{\psi_i(q)}, \qquad \forall q \in V. \tag{6.41}$$

Let ϕ_j be a chart on U_j such that $\phi_j \circ \psi_i^{-1}$ is the identity map on $\psi_i(U_i \cap U_j)$ (such a chart must exist). Let $(\tilde{s}_j, \tilde{D}_j)$ be a parametrization defined by

$$\tilde{s}_j = \hat{s}_j \circ \psi_j \circ \phi_j^{-1} \tag{6.42}$$

where $\widetilde{D}_j = \psi_j \circ \phi_j^{-1}(\widehat{D}_j)$. Since $\psi_j \circ \phi_j^{-1}$ is a diffeomorphism, $\hat{s}_j$ and $\widetilde{s}_j$ are WGO-equivalent. Solve equation (6.42) for $\hat{s}_j$ and substitute into (6.41) to yield

$$(\hat{s}_i)_{\psi_i(q)} \overset{C^r}{=} (\widetilde{s}_j \circ \phi_j \circ \psi_i^{-1})_{\psi_i(q)}, \qquad \forall q \in V. \tag{6.43}$$

Since $\widetilde{s}_j$ is WGO-equivalent to $\hat{s}_j$, and since $\hat{s}_i$ is WGO-equivalent to itself, by Lemma 6.3, $\widetilde{s}_j$ and $\hat{s}_i$ belong to another spline $\widetilde{S}'$ on $(\mathcal{P}, \Phi)$. By construction, $\phi_j \circ \psi_i^{-1}$ is the identity map, so by Definition 6.3, $\hat{s}_i$ and $\widetilde{s}_j$ meet with C^r continuity on V, thus completing the proof of necessity for WG^r continuity.

The proof of necessity of G^r continuity is identical to the proof above, except it must be shown that ψ_j and ϕ_j are coherently oriented, given that ψ_i and ψ_j are coherently oriented. This can be done by letting q be any point in V, and noting that $J[\psi_i \circ \psi_j^{-1}](\psi_i(q))$ is positive by assumption. The following derivation establishes the coherent orientation of ϕ_j and ψ_j:

$$\begin{aligned} J[\phi_j \circ \psi_j^{-1}](\psi_i(q)) &= J[(\psi_i \circ \psi_i^{-1}) \circ \phi_j \circ \psi_j^{-1}](\psi_i(q)) \\ &= J[\psi_i \circ \psi_j^{-1}](\psi_i(q)) \\ &> 0, \end{aligned} \tag{6.44}$$

The second line in equation (6.44) follows from the first line because $\psi_i^{-1} \circ \phi_j$ is the identity by construction.

To prove sufficiency, assume that $\widetilde{s}_i$ and $\widetilde{s}_j$ belong to a spline $\widetilde{S}' = (\{(\widetilde{s}_i, \widetilde{D}_i)\}, C)$ such that $\widetilde{s}_i$ and $\widetilde{s}_j$ meet with C^r continuity on V, and that $\widetilde{s}_i$ and $\widetilde{s}_j$ are WGO-equivalent to $\hat{s}_i$ and $\hat{s}_j$, respectively. Let s_i and s_j be liftings of $\widetilde{s}_i$ and $\widetilde{s}_j$ with respect to which they meet with C^r continuity, and let ϕ_i and ϕ_j be the corresponding charts. That is,

$$\begin{aligned} \widetilde{s}_i &= s_i \circ \phi_i^{-1} \\ \widetilde{s}_j &= s_j \circ \phi_j^{-1} \end{aligned} \tag{6.45}$$

where

$$(\widetilde{s}_i)_{\phi_i(q)} \overset{C^r}{=} (\widetilde{s}_j \circ \phi_j \circ \phi_i^{-1})_{\phi_i(q)}, \qquad \forall q \in V \tag{6.46}$$

and $J[\phi_j \circ \phi_i^{-1}](\phi_i(q)) = 1$. WGO-equivalence implies the existence of diffeomorphisms d_i and d_j such that

$$\begin{aligned} \widetilde{s}_i &= \hat{s}_i \circ d_i \\ \widetilde{s}_j &= \hat{s}_j \circ d_j. \end{aligned} \tag{6.47}$$

Let ψ_i and ψ_j be new charts on U_i and U_j defined by

$$\begin{aligned} \psi_i &= d_i \circ \phi_i \\ \psi_j &= d_j \circ \phi_j. \end{aligned} \tag{6.48}$$

With these definitions, s_i and s_j are liftings of $\hat{s}_i$ and $\hat{s}_j$ relative to (U_i, ψ_i) and (U_j, ψ_j). That is,

$$\begin{aligned} \hat{s}_i &= s_i \circ \psi_i^{-1} \\ \hat{s}_j &= s_j \circ \psi_j^{-1}. \end{aligned} \tag{6.49}$$

Write equation (6.46) in terms of s_i and s_j using equation (6.45), then use Claim 6.1 and Lemma 6.1 to obtain

$$(\hat{s}_i)_{\psi_i(q)} \overset{C^r}{=} (\hat{s}_j \circ \psi_j \circ \psi_i^{-1})_{\psi_i(q)}, \qquad \forall q \in V, \tag{6.50}$$

showing that $\hat{s}_i$ and $\hat{s}_j$ meet with WG^r continuity on V. Moreover, if $\tilde{s}_i$ and $\tilde{s}_j$ are GO-equivalent to $\hat{s}_i$ and $\hat{s}_j$, then $Jd_i > 0$, $Jd_j > 0$, and $J[\phi_j \circ \phi_i^{-1}] = 1$, implying

$$\begin{aligned} J[\psi_j \circ \psi_i^{-1}] &= J[(d_j \circ \phi_j) \circ (d_i \circ \phi_i)^{-1}] \\ &= J[d_j \circ \phi_j \circ \phi_i^{-1} \circ d_i^{-1}] \\ &= Jd_j \quad J[\phi_j \circ \phi_i^{-1}] \quad Jd_j^{-1} \\ &> 0. \end{aligned} \tag{6.51}$$

Thus, ψ_i and ψ_j are coherently oriented, meaning that $\hat{s}_i$ and $\hat{s}_j$ meet with G^r continuity on V. ∎

It is natural to ask what kind of manifolds allow G^r splines to be constructed on them. The next lemma shows that G^r splines can only be constructed on orientable manifolds.

Lemma 6.4: *If $\tilde{S}$ is a G^r spline on $(\mathcal{P}, \Phi)$, then $\mathrm{Int}(\mathcal{P})$ is an orientable manifold. If $\mathrm{Int}(\mathcal{P})$ is not orientable, then it is not possible to construct a G^r spline on $\mathcal{P}$.*

Proof: To prove the first statement, let $\tilde{S}$ be a G^r spline on $(\mathcal{P}, \Phi)$, implying the existence of a connecting basis $\{(U_i, \psi_i)\}$ that is coherently oriented. Since the connecting basis is a basis for $\mathrm{Int}(\mathcal{P})$, $\mathrm{Int}(\mathcal{P})$ must be orientable.

To prove the second statement, assume that $\mathrm{Int}(\mathcal{P})$ is not orientable, and let Φ be an arbitrary tesselation of $\mathcal{P}$. Now, assume there exists a G^r spline $\tilde{S}$ on $(\mathcal{P}, \Phi)$. By the proof of the first statement, $\mathrm{Int}(\mathcal{P})$ is orientable, which is a contradiction. ∎

G^r splines provide a natural orientation for the image of the spline. To investigate this, let $\widetilde{S} = (\{(\hat{s}_i, \widehat{D}_i)\}, \mathcal{C})$ be a G^r spline on $(\mathcal{P}, \Phi)$, $r \geq 1$, with $\dim(\mathcal{P}) = p$. Let $\hat{s}_i$ and $\hat{s}_j$ be two parametrizations of $\widetilde{S}$ that abut at some point $q \in \text{Int}(\mathcal{P})$, and let ψ_i and ψ_j be the charts relative to which they meet with G^r continuity. Consider the Jacobian matrix of $\hat{s}_i$ evaluated at q. To evaluate $\hat{s}_i$ at a point q on the manifold, q must be expressed in terms of the coordinate neighborhood (U_i, ψ_i). Thus, the Jacobian matrix of interest is $D\hat{s}_i(\psi_i(q))$. Each of the p columns of this matrix can be thought of as a vector in $\Re^m$, and since $\hat{s}_i$ is assumed to be regular, the columns are linearly independent. The columns considered as vectors therefore span a vector space of dimension p. In fact, the vector space they span is called the *tangent space of the image of* $\hat{s}_i$ *at* $\hat{s}_i(\psi_i(q))$, denoted $T_{\psi_i(q)}(\hat{s}_i)$.

Example 6.5: *An example of a tangent space was discussed in Section 2.5, although it wasn't identified as such. In that section, the partial derivatives* $\mathbf{G}^{(1,0)}(u_\mathbf{p}, v_\mathbf{p})$ *and* $\mathbf{G}^{(0,1)}(u_\mathbf{p}, v_\mathbf{p})$ *were shown to span the tangent plane of* $\mathbf{G}$ *at* $\mathbf{p}$ *(more precisely, they span the tangent plane of the image of* $\mathbf{G}$ *at the point* $\mathbf{p}$*). To see the connection to the defintion of a tangent space above, note that the components of the vectors* $\mathbf{G}^{(1,0)}$ *and* $\mathbf{G}^{(0,1)}$ *are the columns of the Jacobian matrix* $D\mathbf{G}(u_\mathbf{p}, v_\mathbf{p})$*. For surfaces, the tangent space is two dimensional, and is therefore called the tangent plane. Thus, the columns of* $D\mathbf{G}(u_\mathbf{p}, v_\mathbf{p})$ *span the tangent space* $T_\mathbf{p}(\mathbf{G})$. •

Prior to Example 6.5, we saw that the columns of $D\hat{s}_i(\psi_i(q))$ span $T_{\psi_i(q)}(\hat{s}_i)$. Similarly the columns of the Jacobian matrix of $\hat{s}_j$ span $T_{\psi_j(q)}(\hat{s}_j)$. Recall that $\hat{s}_i$ and $\hat{s}_j$ meet with G^r continuity at q with respect to ψ_i and ψ_j, implying that the first order partials of $\hat{s}_i$ at q and the first order partials of $\hat{s}_j \circ \psi_j \circ \psi_i^{-1}$ at q agree. This condition can be written in matrix form as

$$D\hat{s}_i(\psi_i(q)) = D[\hat{s}_j \circ \psi_j \circ \psi_i^{-1}](\psi_i(q)) \tag{6.52}$$

which by the chain rule (Theorem 6.1) can be written as

$$D\hat{s}_i(\psi_i(q)) = D\hat{s}_j(\psi_j(q)) \cdot D[\psi_j \circ \psi_i^{-1}](\psi_i(q)). \tag{6.53}$$

The matrices $D\hat{s}_i(\psi_i(q)$ and $D\hat{s}_j(\psi_j(q))$ represent bases for $T_{\psi_i(q)}(\hat{s}_i)$ and $T_{\psi_j(q)}(\hat{s}_j)$, respectively. Equation (6.53) represents a change of basis from $D\hat{s}_j(\psi_j(q))$ to $D\hat{s}_i(\psi_i(q))$, with $D[\psi_j \circ \psi_i^{-1}](\psi_i(q))$ acting as the transformation matrix. Thus, $T_{\psi_i(q)}(\hat{s}_i)$ and $T_{\psi_j(q)}(\hat{s}_j)$ are the same space. Moreover, since $\hat{s}_i$ and

$\hat{s}_j$ meet with G^r continuity at q, $D[\psi_j \circ \psi_i^{-1}](\psi_i(q))$ must be orientation preserving. Equation (6.53) therefore shows that there is continuity of *oriented tangent spaces* when parametrizations meet with G^r continuity, $r \geq 1$. Given a G^r spline, we can assign a unique orientation to each point of the image of the spline; namely, the orientation for the tangent space provided by the parametrizations at each point.

Remark 6.13: *Figure 6.12 shows a plot of the Beta-spline basis function. The basis function has first and second derivative discontinuities dictated by equations (3.21) and (3.22) of Chapter 3.*

Figure 6.12 is actually somewhat misleading. Each of the basis segments is parametrized on its own (separate) domain, the domains of adjacent segments being related on the manifold by a transition function. For the case of Figure 6.12, the manifold is $[0, 4]$. *To be completely correct then, before plotting each basis segment over the manifold, the segment should be deformed according to the transition function. However, Figure 6.12 was created by naively plotting each segment on its own domain, then translating the plot of each segment down the u-axis. This corresponds to using a transition function that does not deform the domain. If one were to appropriately deform the domains before plotting, the graph would appear smooth. We will occasionally refer to the "derivative discontinuities" that the Beta constraints introduce, but it should be remembered that these discontinuities are only an artifact of inconsistency.*

6.9. Beta Constraints: Application of the Theory

In terms of the manifold approach to spline construction, the Beta constraints can be viewed as the conditions that result from equation (6.29). The chain rule is used to expand the right side of equation (6.31) in terms of derivatives of $\hat{s}_j$ and derivatives of the transition function $\psi_j \circ \psi_i^{-1}$.

To examine this further, let $\widetilde{S}$ be a spline on a tesselated p-manifold, and let $\hat{s}_1$ and $\hat{s}_2$ be two parametrizations of $\widetilde{S}$ meeting with WG^r continuity on $B = D_1 \cap D_2$. Then there exist connecting neighborhoods (U_1, ψ_1) and (U_2, ψ_2) such that

$$(\hat{s}_1)_{\psi_1(q)} \overset{C^r}{=} \left(\hat{s}_2 \circ \psi_2 \circ \psi_1^{-1}\right)_{\psi_1(q)}, \qquad \forall q \in B. \tag{6.54}$$

In Chapter 2, the notation $\mathcal{CR}_{i,j}$ was used to denote the function resulting from a chain rule expansion for the $(i,j)^{\text{th}}$ partial of its arguments. We extend this notation to the case of $\mathcal{CR}_{\vec{k}}$, for $\vec{k}$ a general p-tuple. That is, if $\widetilde{f} = \hat{f} \circ d$,

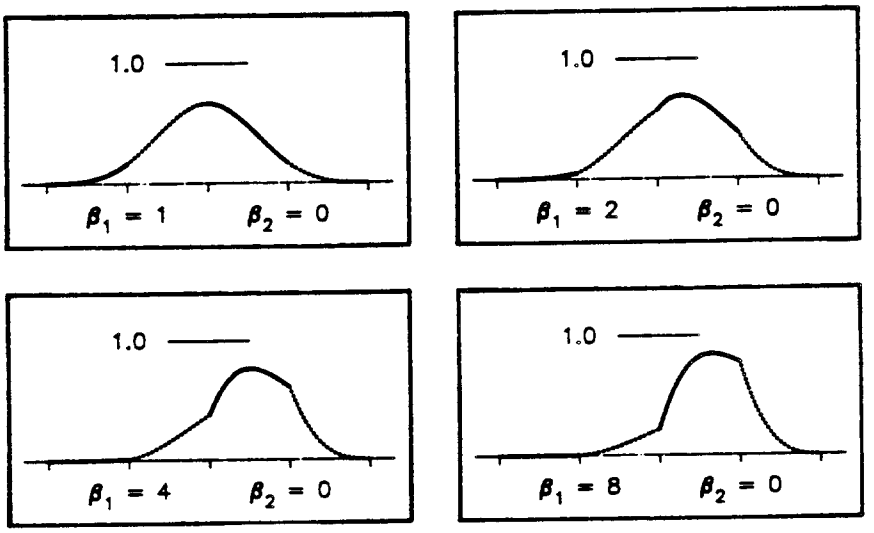

Figure 6.12. *A graph of the Beta-spline basis segments comprising the Beta-spline blending function. The shape parameter values are as shown in the individual figures (from Barsky & Beatty [5]).*

$d : U \to V$, $f : V \to \Re^p$, then

$$\tilde{f}^{\vec{k}}(x) = \mathcal{CR}_{\vec{k}}(\hat{f}^{\vec{t}}(d(x)), d^{\vec{t}}(x)) \tag{6.55}$$

for all $\vec{t} = (t_1, ..., t_p)$ such that $|\vec{t}| \leq |\vec{k}|$. The tuple $\vec{t}$ is merely notational shorthand for reducing the number of arguments to $\mathcal{CR}_{\vec{k}}$ that must be explicitly written out. With this notation, equation (6.54) can be rewritten as the set of constraints:

$$\hat{s}_1^{\vec{k}}(\psi_1(q)) = \mathcal{CR}_{\vec{k}}(\hat{s}_2^{\vec{t}}(\psi_2(q)), (\psi_2 \circ \psi_1^{-1})^{\vec{t}}(\psi_1(q))), \qquad \forall q \in B, \tag{6.56}$$

$$1 \leq |\vec{t}| \leq |\vec{k}| \leq r.$$

In Section 2.5.3, it was demonstrated that not all constraints implied by equation (2.36) were independent. We are faced with the same situation here. To show which constraints are dependent, we make some simplifying assumptions. The assumptions will be relaxed later in the section. In particular, we will initially assume that:

- The boundary set B is a manifold of dimension $p - 1$. Thus, for curves ($p = 1$), the boundary between $\hat{s}_1$ and $\hat{s}_2$ is a point, and for surfaces ($p = 2$), the boundary between $\hat{s}_1$ and $\hat{s}_2$ is a curve.
- The parametrization $\hat{s}_1(x_1, ..., x_p)$ is such that the image set $\hat{s}_1(\psi_1(B))$ is generated by holding x_1 fixed, say at 0, and letting the other $p - 1$

variables take on all possible values. Stated alternately, if $q \in B$, then $\psi_1(q) = (0, q_2, ..., q_p)$.

With these assumptions, it is straightforward to show that constraints (6.56) hold if and only if the constraints corresponding to $\vec{k} = (k_1, 0, ..., 0)$ hold for $k_1 = 1, ..., r$. All other constraints implied by (6.56) are consequences of these r constraints. One can then show that only the derivatives $(\psi_2 \circ \psi_1^{-1})^{(t_1,...,0)}$, $t_1 = 1, ..., r$, appear in the r constraints. It is therefore convenient to introduce functions $\beta_i : \psi_1(B) \subset \Re^{p-1} \to \Re^p$, $i = 0, 1, ..., r$, defined by

$$\beta_i = (\psi_2 \circ \psi_1^{-1})^{(i,0,...,0)}|B. \tag{6.57}$$

Note that each of the β's is a $p-1$ variate function into $\Re^p$, and the collection β_i, $i = 0, ..., r$ locally characterizes the transition function to order r. Here, "locally" refers to an open set in $\psi_1(U_1)$ containing $\psi_1(B)$. To explicitly show that the β's are lower dimension restrictions of derivatives of the transition function, it is convenient to introduce a projection operator $P_1 : \Re^p \to \Re^{p-1}$ defined by

$$P_1 \circ \psi_1(q) = P_1(q_1, q_2, ..., q_p) = (q_2, ..., q_p). \tag{6.58}$$

Equation (6.57) can then be written as

$$\beta_i(P_1 \circ \psi_1(q)) = (\psi_2 \circ \psi_1^{-1})^{(i,0,...,0)}(\psi_1(q)), \qquad q \in B. \tag{6.59}$$

It is important to realize that β_0 is special in that it is fixed by the C^0 assumption, and hence, by the equivalence class $\mathcal{C}$. That is, if ψ_1' and ψ_2' are charts similar to ψ_1 and ψ_2, then $\beta_0' = \beta_0$, but the functions $\beta_1', ..., \beta_i'$ will in general differ from $\beta_1, ..., \beta_i$. Reffering back to Chapter 2, we can identify the $\beta_1, ..., \beta_i$ as *shape parameters*, collectively referred to as a *shape set*. The fact that the transition function has a non-singular Jacobian matrix implies that the β's satisfy

$$\begin{aligned} J[(\psi_2 \circ \psi_1^{-1})](\psi_1(q)) &= \det[\beta_1(P_1 \circ \psi_1(q))\ \beta_0^{\vec{j}_1}(P_1 \circ \psi_1(q)) \cdots \beta_0^{\vec{j}_{p-1}}(P_1 \circ \psi_1(q))] \\ &\neq 0, \end{aligned} \tag{6.60}$$

where $\vec{j}_s$ is a $p-1$ tuple as described in Section 1.2.

With these definitions and observations, $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r continuity on B if and only if

$$\hat{s}_1^{(k_1,0,...,0)}(\psi_1(q)) = \mathcal{CR}_{(k_1,0,...,0)}(\hat{s}_2^{\vec{t}}(\psi_2(q)),\ \beta_i(P_1 \circ \psi_1(q))), \qquad \forall q \in B, \tag{6.61}$$

$$1 \leq i, |\vec{t}| \leq k_1 \leq r.$$

Remark 6.14: *The reader may notice that equation (6.61) is burdensome notation for simple partial differentiation. We have chosen to obfuscate the discussion here in an effort to make a later generalization more transparent.*

If $\hat{s}_1$ and $\hat{s}_2$ meet with the stronger condition of G^r continuity on B, then constraints (6.61) hold, and the coherent orientation of ψ_1 and ψ_2 implies that the β's satisfy

$$\det[\beta_1(P_1 \circ \psi_1(q))\ \beta_0^{\vec{j}_1}(P_1 \circ \psi_1(q))\ \cdots\ \beta_0^{\vec{j}_p - 1}(P_1 \circ \psi_1(q))] > 0. \tag{6.62}$$

The constraints implied by equation (6.61) are the *p-variate Beta constraints* for the case where the boundary is of dimension $p-1$. The argument given above shows the necessity of the p-variate Beta constraints. We now state the result of primary practical importance — the necessity *and* sufficiency of the p-variate Beta constraints for (weak) geometric continuity of the spline. In practice, this result means that the functions $\beta_1, ..., \beta_r$ can be arbitrarily chosen, subject to equation (6.60).

The crux of the sufficiency proof lies in the ability to find a transition function whose derivatives match the β's when evaluated along the boundary. In other words, given the β's defined along the boundary, we wish to find a diffeomorphism, defined on a higher dimension, that has the β's as a restriction to the boundary. The following extension lemma details when such a diffeomorphic extension exists.

Lemma 6.5: *Let $f_{\vec{i}}(y) : \Re^b \to \Re^p$, $y = (y_1, ..., y_b)$, $\vec{i} = (i_1, ..., i_n)$, be a collection of C^∞ functions defined on a neighborhood V of $0 \in \Re^b$. Then there exists a C^∞ map $F(x, y) : \Re^n \times \Re^b \to \Re^p$, $x = (x_1, ..., x_n)$, defined on a neighborhood of $0 \in \Re^n \times \Re^b$ such that*

$$\frac{\partial^{|\vec{i}|} F}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}}(0, y) = f_{\vec{i}}(y), \tag{6.63}$$

for all $\vec{i}$ such that $1 \leq |\vec{i}| \leq r$. Moreover, if $f_{(0,...,0)}(y)$ is 1-1 on a compact subset W of V, $n + b = p$, and

$$\det[f_{\vec{j}_1}(y)\ \cdots\ f_{\vec{j}_n}(y)\ \frac{\partial f_{(0,...,0)}}{\partial y_1}(y)\ \cdots\ \frac{\partial f_{(0,...,0)}}{\partial y_b}(y)] \neq 0 \tag{6.64}$$

$\forall y \in W$, then there is a neighborhood of W in $\Re^n \times \Re^b$ on which F is a diffeomorphism.

Proof: An example of a function satisfying the requirements of the lemma is

$$F(x,y) = \sum_{|\vec{i}|=0}^{r} \frac{x_1^{i_1} \cdots x_n^{i_n}}{i_1! \cdots i_n!} f_{\vec{i}}(y). \tag{6.65}$$

If $f_{(0,\ldots,0)}(y)$ is 1-1 on W, then F is 1-1 on W, and if the f's satisfy equation (6.64), then F is regular on W. The generalized Inverse Function Theorem (Theorem 6.4) then guarantees that there is a neighborhood of W in $\Re^n \times \Re^b$ on which F is a diffeomorphism. ∎

Theorem 6.11: *Let $\widetilde{S} = (\{(\hat{s}_i, \hat{D}_i)\}, C)$ be a regular parametric spline on $(\mathcal{P}, \Phi)$, $\dim(\mathcal{P}) = p$, let S given by $\{s_i, D_i\}$ be a lifting relative to $\{(U_i, \psi_i)\}$, and let $\hat{s}_1$, $\hat{s}_2$ abut on a boundary $B = D_1 \cap D_2$ of dimension $p-1$. $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r continuity on B if and only if there exists a shape set β_i, $i = 1, \ldots, r$, satisfying constraints (6.61) and (6.60) where $\beta_0(P_1 \circ \psi_1(q)) = (\psi_2 \circ \psi_1^{-1})(\psi_1(q))$, for all $q \in B$. Moreover, $\hat{s}_1$ and $\hat{s}_2$ meet with G^r continuity on B if and only if there exists a shape set satisfying constraints (6.61) and (6.62).*

Proof: The discussion preceding this theorem shows that if $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r or G^r continuity, then there exists a shape set that satisfies the conditions of the theorem.

To prove the converse, assume that a shape set exists satisfying the conditions of the theorem for WG^r continuity, and let $\beta_0(P_1 \circ \psi_1(q)) = (\psi_2 \circ \psi_1^{-1})(\psi_1(q))$. By setting $n = 1$, $b = p-1$, and $W = B$ in Lemma 6.5, the lemma can be used to show that there exists a diffeomorphism $d : \psi_1(U_1 \cap U_2) \to \psi_2(U_1 \cap U_2)$ such that

$$d^{(i,0,\ldots,0)}(\psi_1(q)) = \beta_i(P_1 \circ \psi_1(q)), \quad \forall q \in B,\ i = 0, \ldots, r. \tag{6.66}$$

From equation (6.66), we note that every diffeomorphism satisfying equation (6.66) also satisfies

$$\begin{aligned} d(\psi_1(q)) &= \beta_0(P_1 \circ \psi_1(q)) \\ &= (\psi_2 \circ \psi_1^{-1})(\psi_1(q)) \\ &= \psi_2(q). \end{aligned} \tag{6.67}$$

Since d is a diffeomorphism on $\psi_1(U_1 \cap U_2)$, we may choose a chart $\phi_2 : U_2 \to \psi_2(U_2)$ such that

$$d = \phi_2 \circ \psi_1^{-1}. \tag{6.68}$$

Moreover, ϕ_2 is similar to ψ_2 since if $q \in B$, then from equation (6.68), $\phi_2(q) = d \circ \psi_1(q) = d(\psi_1(q))$, which by equation (6.67) implies that $\phi_2(q) = \psi_2(q)$ for all $q \in D_1 \cap D_2$.

Having chosen a chart ϕ_2 similar to ψ_2 and satisfying equation (6.68), the steps leading from equation (6.54) to equation (6.61) may be reversed, using ϕ_2 in place of ψ_2, yielding

$$(\hat{s}_1)_{\psi_1(q)} \overset{C^r}{=} (\hat{s}_2 \circ \phi_2 \circ \psi_1^{-1})_{\psi_1(q)}, \quad \forall q \in B. \tag{6.69}$$

Equation (6.69) can be written in terms of s_1 and s_2 as

$$\begin{aligned}(s_1 \circ \psi_1^{-1})_{\psi_1(q)} &\overset{C^r}{=} (s_2 \circ \psi_2^{-1} \circ \phi_2 \circ \psi_1^{-1})_{\psi_1(q)} \\ &\overset{C^r}{=} (s_2' \circ \psi_1^{-1})_{\psi_1(q)}\end{aligned} \tag{6.70}$$

$\forall q \in B$, where $s_2' = s_2 \circ \psi_2^{-1} \circ \phi_2$. Since ϕ_2 was chosen to be similar to ψ_2, s_2' is similar to s_2. Therefore, $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r continuity on B with respect to a lifting $S' \in \mathcal{C}$, where S' is identical to S except that s_2 is replaced with s_2'.

If the β's are known to satisfy equation (6.62), then d is an orientation preserving diffeomorphism, implying that ϕ_2 and ψ_1 are coherently oriented. Equation (6.69), Lemma 6.1, and Definition 6.11 then imply that $\hat{s}_1$ and $\hat{s}_2$ meet with G^r continuity on B. ∎

Remark 6.15: *Recall that geometric continuity reduces to parametric continuity when the abutting parametrizations share the same local coordinate system. This fact is reflected in the Beta constraints in that the Beta constraints reduce to requiring continuity of derivatives when the transition function has a Jacobian matrix equal to the identity matrix. This in turn implies that* $\beta_0(q_2, ..., q_p) = (c_1, y_2, ..., y_p)$ *and* $\beta_1(q_2, ..., q_p) = (1, 0, ..., 0)$, *where* c_1 *is a constant, the y's differ from the q's by at most a constant, and* $\beta_2, ..., \beta_r$ *are all equal to the zero function.*

Theorem 6.11 exhibits the conditions that must be satisfied between two p-variate parametrizations that abut along a boundary whose dimension is $p-1$. The behavior of the transition function along the boundary is completely specified by the C^0 assumption, the only freedom being in the cross boundary dimension. Thus, the only independent constraints in (6.61) are those corresponding to the cross boundary dimension. For the situation covered by Theorem 6.11, there is only one cross boundary dimension, and therefore only r constraints for r^{th} order continuity, with the functions $\beta_1, ..., \beta_r$ determining the behavior of the transition function along that direction. If the boundary was of dimension $p-2$, then there would be 2 cross boundary dimensions, and the shape set would consist of doubly

subscripted $(p-2)$-variate functions. That is, the shape set would be given by $\beta_{i,j} : \Re^{p-2} \rightarrow \Re^p$, for $1 \leq i+j \leq r$, and the independent constraints would be

$$\hat{s}_1^{(k_1,k_2,0,\ldots,0)}(\psi_1(q)) = \mathcal{CR}_{(k_1,k_2,0,\ldots,0)}(\hat{s}^{\vec{t}}(\psi_2(q)), \beta_{i,j}), \tag{6.71}$$

$$1 \leq i+j,\ |\vec{t}| \leq k_1 + k_2 \leq r,$$

assuming that the boundary is generated by fixing the first two variables of $\hat{s}_1$, letting the other $p-2$ vary. Equation (6.71) constitutes $r(r+3)/2$ constraints, and therefore introduces $r(r+3)/2$ shape parameters.

Example 6.6: *As a specific example, consider two surfaces ($p = 2$) meeting at a point with G^2 continuity. For this case, there are $2(2+3) = 10$ shape parameters, each a 0-variate function into $\Re^2$, i.e., a two component vector of real numbers.* •

In the general case, two p-variate parametrizations can meet along a boundary of dimension b. If we assume that the boundary is generated by holding the first b variables of $\hat{s}_1$ fixed, then the shape set consists of $\beta_{\vec{i}} : \Re^b \rightarrow \Re^p$, $\vec{i} \in \mathbf{Z}_+^b$, $1 \leq |\vec{i}| \leq r$, with the independent constraints being given by

$$\hat{s}_1^{\vec{k}}(\psi_1(q)) = \mathcal{CR}_{\vec{k}}(\hat{s}^{\vec{t}}(\psi_2(q)),\ \beta_{\vec{i}}), \qquad \forall q \in B, \tag{6.72}$$

$$1 \leq |\vec{i}|,\ |\vec{t}| \leq |\vec{k}| \leq r$$

where $\vec{k} = (k_1, \ldots, k_{p-b}, 0, \ldots, 0) \in \mathbf{Z}_+^p$, $\vec{t} \in \mathbf{Z}_+^p$, $\vec{i} \in \mathbf{Z}_+^b$, and $\beta_0 = \psi_2 \circ \psi_1^{-1}|B$. The shape parameters must satisfy a rather ugly non-singularity condition similar to (6.64). Specifically,

$$\det[\beta_{\vec{j}_1} \cdots \beta_{\vec{j}_{p-b}}\ \beta_0^{\vec{j}_1} \cdots \beta_0^{\vec{j}_b}] \neq 0. \tag{6.73}$$

The general theorem of necessity and sufficiency of the Beta constraints for two parametrizations may be stated as:

Theorem 6.12: *Let $\widetilde{S} = (\{(\hat{s}_i, \widehat{D}_i)\}, C)$ be a regular parametric spline on $(\mathcal{P}, \Phi)$, $\dim(\mathcal{P}) = p$, and let $\hat{s}_1$, $\hat{s}_2$ abut on a boundary $B = D_1 \cap D_2$ of dimension b. $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r continuity on B if and only if there exists a shape set $\beta_{\vec{i}} : \psi_1(B) \rightarrow \psi_2(U_1 \cap U_2)$, $\vec{i} \in \mathbf{Z}_+^b$, $|\vec{i}| = 1, \ldots, r$, satisfying constraints (6.72) and (6.73) where $\beta_0 = \psi_2 \circ \psi_1^{-1}|B$. Moreover, $\hat{s}_1$ and $\hat{s}_2$ meet with G^r continuity on B if and only if there exists a shape set satisfying constraints (6.72) where the expression in (6.73) is positive.*

Proof: The proof follows an analysis similar to the proof of Theorem 6.12, except that the form of the local characterization of the diffeomorphism d must account for the increased number of cross boundary directions. Lemma 6.5 may still be used, this time with $n = p - b$. ∎

In Chapter 2, heuristic rules for deriving the Beta constraints were given for curves, and surfaces. Similar heuristics can be derived in a straightforward manner for the determination of the p-variate Beta constraints given in symbolic form in equation (6.72).

Theorem 6.12 treats the case where only two parametrizations abut at a point, but in general there can be any number of parametrizations abutting at a point. Theorem 6.12 is still valid in that there must exists a shape set satisfying constraints (6.72) between every pair of parametrizations abutting at the point. Thus, if $\hat{s}_1$, $\hat{s}_2$, and $\hat{s}_3$ are three parametrizations abutting at q, then there must be a shape set relating $\hat{s}_1$ and $\hat{s}_2$, another relating $\hat{s}_2$ and $\hat{s}_3$, and still another relating $\hat{s}_1$ and $\hat{s}_3$. However, the shape sets must be ***compatible*** in the sense that if the shape sets between $\hat{s}_1$ and $\hat{s}_2$, and $\hat{s}_2$ and $\hat{s}_3$ are specified, then the shape set between $\hat{s}_1$ and $\hat{s}_3$ is uniquely determined at q.

6.9.1. Transition Graphs

To examine further the question of compatibility between shape sets, it is convenient to introduce the notion of a ***transition graph at a point.*** Let $q \in \text{Int}(\mathcal{P})$ be a point where A parametrizations $\hat{s}_1, ..., \hat{s}_A$ abut with geometric continuity with respect to the connecting neighborhoods $(U_1, \psi_1), ..., (U_A, \psi_A)$, and let τ_{ji} denote the transition function from $\hat{s}_i$'s coordinate system to $\hat{s}_j$'s coordinate system. That is,

$$\tau_{ji} = \psi_j \circ \psi_i^{-1}. \tag{6.74}$$

The transition graph at q is an ***undirected graph*** (in the computer science sense of a graph), containing A nodes, one for each parametrization. Label each node with the index of the parametrization to which it corresponds. For each pair of nodes (i, j) for which a shape set is specified, add an arc between node i and node j, labeling the arc by τ_{ji}. Strictly speaking, the arc should be a directed arc from i to j and labeled by τ_{ji}. However, since τ_{ji} is a diffeomorphism, the reverse arc corresponding to the transition from j to i is labeled with $\tau_{ij} = \tau_{ji}^{-1}$. Thus, the label on a directed arc uniquely determines the label on the reverse arc. Similarly,

a shape set from $\hat{s}_i$ to $\hat{s}_j$ uniquely determines the inverse shape set from $\hat{s}_j$ to $\hat{s}_i$. In this way, the arcs can be considered to be undirected.

Suppose there is a path from node i to node k in the transition graph for a point q. The transition function τ_{ki} from $\hat{s}_i$ to $\hat{s}_k$ is the composition of the labels of the arcs on the path from i to k. This transition function must be unique, meaning that the path from i to k must be unique. Since the Beta constraints must hold between every pair of parametrizations, there must be a unique path from every node in the transition graph to every other node. Thus, the transition graph must be a *minimal spanning graph*, otherwise known as a *spanning tree* (cf. Aho et al [1] for an introduction to graph theoretic concepts).

A spanning tree with A nodes has $A-1$ arcs, so there are $A-1$ freely selectable shape sets at a point where A parametrizations abut. At such a point, there are $A(A-1)/2$ possible shape sets, one for each pair of abutting parametrizations. Of these, the $A-1$ freely selected shape sets must be chosen so that the transition graph forms a spanning tree.

Example 6.7: *The notion of a transition graph for curves is trivial since there can be at most two nodes in the graph for a point, and therefore only one arc. As an example of the process of shape set selection for surfaces, consider the situation shown in Figure 6.13 where four parametrizations $\hat{s}_1, ..., \hat{s}_4$ abut at q on a 2-manifold. There are many possible transition graphs for q, several of which are shown in Figure 6.13. In (d) for instance, a shape set is specified between $\hat{s}_1$ and $\hat{s}_2$, and one between $\hat{s}_1$ and $\hat{s}_4$. The shape set between $\hat{s}_2$ and $\hat{s}_4$ is therefore completely determined by $\tau_{41} \circ \tau_{12}$, the composite transition function corresponding to the path from 2 to 4.* •

6.10. Equivalence Theorems

There are many equivalent ways to characterize (weak) geometric continuity, several of which have already been presented. Chapter 2 based a characterization on the existence of GO-equivalent parametrizations. Theorems 6.8, 6.9 and 6.12 provide additional characterizations. In this section, we present two more equivalence theorems: one based on the notion of smooth curves on splines, and one based on covariant differentiation.

In what follows, let $\widetilde{S} = (\{(\hat{s}_i, \widehat{D}_i)\}, C)$ be a parametric spline on $(\mathcal{P}, \Phi)$, $\dim(\mathcal{P}) > 1$, and let $(\hat{s}_1, \widehat{D}_1)$ and $(\hat{s}_2, \widehat{D}_2)$ be the only two parametrizations abutting at a point $q \in D_1 \cap D_2$. It is important to realize that we are establishing

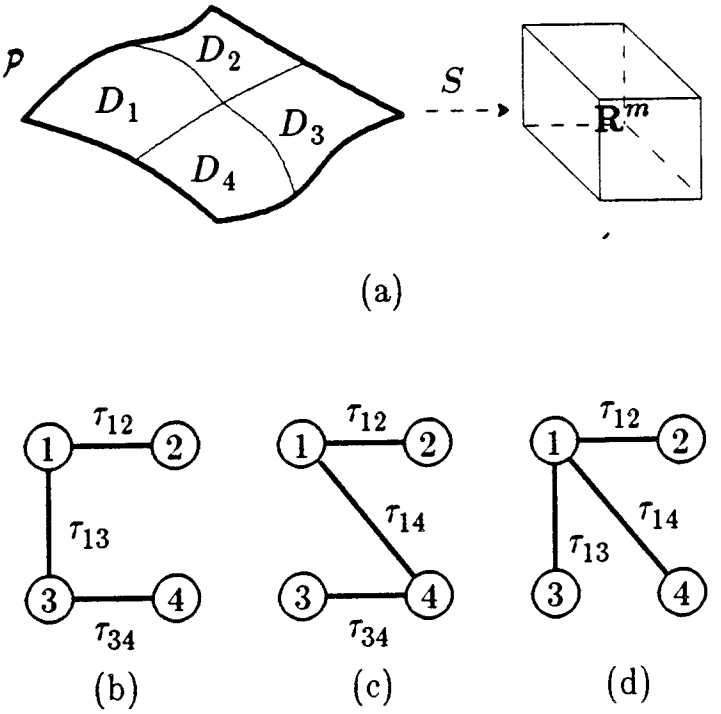

Figure 6.13. *In Figure (a) above is shown a point q on a surface. Figures (b), (c), and (d) represent several of the possible transition graphs for q.*

equivalent definitions of WG^r continuity *at a point,* not on a set. If the conjecture referred to in Remark 6.10 is correct, then continuity on a set follows immediately. However, it must be emphasized that in the absence of the conjecture the results of this section hold only at a point.

Let us first examine the smoothness of curves on the image of the spline. Intuitively, if two parametrizations meet in such a way that all smooth curves on one can be smoothly extended onto the other, then parametrizations must be smoothly in a geometric sense. Note that this is not the case in Figure 2.16 of Chapter 2, even though the patches have a common tangent plane. The idea of smooth extension is formalized in the following equivalence theorem.

Theorem 6.13: *Let $\alpha(t) : (-1, 1) \to P$ be a regular C^∞ curve on the manifold P such that $\alpha(0) = q$. Then $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r continuity at q if and only if there is an abstract spline $S \in \mathcal{C}$ such that $S \circ \alpha$ is regular and C^r at 0 for all α.*

Proof: If $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r continuity at q, then by 6.8, there is an abstract

spline $S \in \mathcal{C}$ that is an immersion at q, implying that the curve $S \circ \alpha$ on $S(\text{Int}(P))$ is regular and C^r at 0.

To prove the converse, let $S \in \mathcal{C}$ be such that $S \circ \alpha$ is regular and C^r at 0, and let s_1 and s_2 belong to S where s_1 is a lifting of $\hat{s}_1$ relative to (U_1, ψ_1), and s_2 is a lifting of $\hat{s}_2$ relative to (U_2, ψ_2). The fact that $S \circ \alpha$ is regular and C^r at 0 implies that

$$(s_1 \circ \alpha)^{(i)}(0) = (s_2 \circ \alpha)^{(i)}(0), \qquad i = 0, ..., r, \tag{6.75}$$

for all α. If equation (6.75) is true for all α, then it is true for all α such that the image of α in $\hat{D}_1$ is linear. That is, equation (6.75) is assumed to hold for all α_1 such that

$$\alpha_1(t) = \psi_1 \circ \alpha(t) = \psi_1(q) + \vec{v}t \tag{6.76}$$

where $\vec{v} = (v_1, ..., v_p) \in \Re^p$. By letting $\hat{s}_1 = s_1 \circ \psi_1^{-1}$, $\hat{s}_2 = s_2 \circ \psi_2^{-1}$, and $\alpha_1 = \psi_1 \circ \alpha$, equation (6.75) may be rewritten as

$$(\hat{s}_1 \circ \alpha_1)^{(i)}(0) = (\hat{s}_2 \circ \psi_2 \circ \psi_1^{-1} \circ \alpha_1)^{(i)}(0), \qquad i = 0, ..., r, \tag{6.77}$$

By setting i to one in equation (6.77), and using the chain rule, we find that

$$D\hat{s}_1(\psi_1(q)) \cdot D\alpha_1(0) = D[\hat{s}_2 \circ \psi_2 \circ \psi_1^{-1}](\psi_1(q)) \cdot D\alpha_1(0). \tag{6.78}$$

Since equation (6.78) must hold for all α_1, it must be that

$$D[\hat{s}_1](\psi_1(q)) = D[\hat{s}_2 \circ \psi_2 \circ \psi_1^{-1}](\psi_1(q)). \tag{6.79}$$

Equation (6.79) may be rewritten as

$$(\hat{s}_1)_{\psi_1(q)} \overset{C^1}{=} (\hat{s}_2 \circ \psi_2 \circ \psi_1^{-1})_{\psi_1(q)}, \tag{6.80}$$

which implies that $\hat{s}_1$ and $\hat{s}_2$ meet with WG^1 continuity. Using Theorem 6.10, $\hat{s}_2$ can be reparametrized to obtain $\hat{s}_2'$ that meets $\hat{s}_1$ with C^1 continuity at q. If ψ_2' denotes the chart relative to which s_2 is a lifting of $\hat{s}_2'$, then equation (6.77) can be written as

$$(\hat{s}_1 \circ \alpha_1)^{(i)}(0) = (\hat{s}_2' \circ \psi_2' \circ \psi_1^{-1} \circ \alpha_1)^{(i)}(0), \qquad i = 0, ..., r, \tag{6.81}$$

where $D[\psi_2' \circ \psi_1^{-1}](\psi_1(q))$ is the identity matrix. Using the chain rule and the linear form of α_1, the left and right sides of equation (6.81) can be expanded as

$$\begin{aligned} (\hat{s}_1 \circ \alpha_1)^{(i)}(0) &= \sum_{|\vec{k}|=1}^{i} \hat{s}_1^{\vec{k}}(\psi_1(q)) v_1^{k_1} \cdots v_p^{k_p} \\ (\hat{s}_2' \circ \psi_2' \circ \psi_1^{-1} \circ \alpha_1)^{(i)}(0) &= \sum_{|\vec{k}|=1}^{i} \hat{s}_2'^{\vec{k}}(\psi_2'(q)) v_1^{k_1} \cdots v_p^{k_p} \end{aligned} \tag{6.82}$$

for $i = 0, ..., r$, where $\vec{k} = (k_1, ..., k_p)$. Substitution of equations (6.82) into equation (6.81), followed by rearrangement yields

$$\sum_{|\vec{k}|=1}^{i} \left[\hat{s}_1^{\vec{k}}(\psi_1(q)) - \hat{s}_2'^{\vec{k}}(\psi_2'(q)) \right] v_1^{k_1} \cdots v_p^{k_p} = 0. \tag{6.83}$$

Since equation (6.81) must hold for all α_1, equation (6.83) must hold for all $v_1, ..., v_p$, and can therefore be considered to be a polynomial in $v_1, ..., v_p$. By uniqueness of polynomials, the sum can be zero only if each term is zero, implying that

$$\hat{s}_1^{\vec{k}}(\psi_1(q)) = \hat{s}_2'^{\vec{k}}(\psi_2'(q)),$$

for all $\vec{k}$ such that $1 \leq |\vec{k}| \leq r$, which in turn implies that $\hat{s}_1$ and $\hat{s}_2'$ meet with C^r continuity at q. By Theorem 6.10, $\hat{s}_1$ and $\hat{s}_2$ must meet with WG^r continuity at q. ∎

From an intuitive standpoint, Theorem 6.13 is rather pleasing in that it shows that the definition of weak geometric continuity can be "bootstraped". That is, weak geometric continuity for objects of parametric dimension larger than 1 can be defined in terms of weak geometric continuity for curves. For instance, a loosely worded corollary to Theorem 6.13 is: two surface patches meet with WG^r continuity at q if and only if all curves on the composite surface passing through q are WG^r curves. The same can be said for objects of higher parametric dimension.

The next equivalence theorem is based on the notion of *covariant derivatives* from tensor analysis. Intuitively, covariant derivatives are tensors that capture the differential properties of a manifold. Since they are tensors, a statement of equality is guaranteed to be coordinate independent, and are therefore independent of the parametrizations. The reader is referred to Boothby [14] or Synge & Schild [59] for a complete treatment of covariant differentiation.

Before stating the theorem, there is one technical point that must be addressed. Covariant derivatives, like all tensors, "live" on the tangent space, or a finite cartesian product thereof. Thus, before discussing equality of covariant derivatives, we must be sure that the parametrizations have a common tangent space at the point of interest. This is equivalent to requiring the parametrizations to meet with WG^1 continuity at q. The theorem of interest is may now be stated as:

Theorem 6.14: *Let $\hat{s}_1$ and $\hat{s}_2$ be as above, meeting with WG^1 continuity at q. $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r continuity at q if and only if the first r covariant derivatives are equal at q.*

Proof: If $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r continuity at q, then by Theorem 6.8, there is an abstract spline $S \in \mathcal{C}$ that is a C^r immersion at q. The image of the spline is therefore locally a C^r manifold, implying that covariant derivatives up to order r are continuous [59].

To prove the converse, let s_1 and s_2 be liftings of $\hat{s}_1$ and $\hat{s}_2$, and assume that covariant derivatives up to order r are continuous at q. Express s_1 and s_2 in terms of the *geodesic coordinate system* [59] at q to obtain $\hat{s}'_1$ and $\hat{s}'_2$. In this coordinate system, the covariant derivative reduces to partial differentiation at q, implying that $\hat{s}'_1$ and $\hat{s}'_2$ meet with C^r continuity at q. $\hat{s}'_1$ and $\hat{s}'_2$ are WGO-equivalent to $\hat{s}_1$ and $\hat{s}_2$, respectively, so by Theorem 6.10, $\hat{s}_1$ and $\hat{s}_2$ meet with WG^r continuity at q. ∎

For curves, covariant differentiation is equivalent to differentiation with respect to arc length; thus, as a corollary to Theorem 6.14 we have:

Theorem 6.15: *Two curves segments meet with WG^r continuity at a point if and only if the first r arc length derivatives agree at the point.*

Theorem 6.15 establishes the equivalence between this work, and the previous work of Barsky & DeRose [6].

For surface, the first and second covariant derivatives are equivalent characterizations of the tangent plane and Dupin indicatrix. We therefore have the following as a corollary to Theorem 6.14.

Theorem 6.16: *Two surface patches meet with WG^2 continuity at a point if and only if they have common position, tangent plane, and Dupin indicatrix.*

6.11. Summary

In this chapter, we have taken an inherently parametric view of spline construction by starting with a collection of parametrizations, each defined on its own domain. The parametrizations were then lifted onto a tesselated, differentiable manifold of appropriate dimension and topology to define an equivalence class of abstract splines. The collection of parametrizations, together with an equivalence

class of abstract splines formed the basis for a definition of parametric splines. A parametrization independent measure of continuity called weak geometric continuity was then defined by requiring that the parametric spline represent a C^r immersion. Equivalent characterizations were then identified and proven, many of which are summarized in Table 6.1.

Only the most basic properties of manifolds and immersions have been used to characterize geometric continuity. The application of more powerful manifold theoretic results seems extremely promising. It is hoped that the use of manifold theory in the solution of problems encountered in CAGD will be an active area of future research.

$\hat{s}_1, \hat{s}_2$ meet		Equivalently	Follows From	Remarks
WG^r	$\Longleftrightarrow$	$\hat{s}_1$, $\hat{s}_2$ represent a C^r immersion at q	Thm 6.8	
WG^r	$\Longleftrightarrow$	There exist WGO-equivalent parametrizations that meet with C^r continuity at q	Thm 6.10	
G^r	$\Longleftrightarrow$	There exist GO-equivalent parametrizations that meet with C^r continuity at q	Thm 6.10	
WG^r	$\Longleftrightarrow$	Locally graph of a C^r real-valued function	Thm 6.9	Codimension 1
WG^r	$\Longleftrightarrow$	Covariant derivatives up to order r agree at q	Thm 6.14	
WG^r	$\Longleftrightarrow$	Arc length derivatives up to order r agree at q	Thm 6.15	Curves
WG^2	$\Longleftrightarrow$	Tangent plane, Dupin indicatrix agree at q	Sec 6.8, Thm 6.16	Surfaces
G^2	$\Longleftrightarrow$	Oriented tangent plane and Dupin indicatrix agree at q	Thm 6.14	Surfaces
WG^1	$\Longleftrightarrow$	Tangent spaces agree at q	Sec 6.8	
G^1	$\Longleftrightarrow$	Oriented tangent spaces agree at q	Sec 6.8	
WG^r	$\Longleftrightarrow$	Smooth curves on $\hat{s}_1$ passing through q can be extended with WG^r continuity onto $\hat{s}_2$	Thm 6.13	

Table 6.1: *Equivalent measures of continuity at a point* $q \in \mathrm{Int}(\mathcal{P})$.

7

Conclusions

A parametrization independent measure of continuity called geometric continuity has been defined and characterized in two different, but equivalent, ways. The first characterization was based on the notion of reparametrization. The second approach was based on the theory of differentiable manifolds. The reparametrization approach is the most convenient when building splines; the manifold approach most appropriate for proving statements about the nature of geometric continuity.

The basic definition of geometric continuity is not useful in practice, so a set of necessary and sufficient conditions, called the Beta constraints, were derived from the basic definition. The Beta constraints were shown to result from a straightforward application of the chain rule: the univariate chain rule for curves, the bivariate chain rule for surfaces, the trivariate chain rule for volumes, and so on for splines of arbitrary parametric dimension.

The Beta constraints provide for the introduction of shape parameters or shape functions that can be used to modify the shape of the spline, independent of other controls a designer has over shape. It was shown that for curves, n shape parameters are introduced at each joint that is stitched together with G^n continuity. For surfaces $2n$ shape functions defined along the boundary curve are introduced when two surface patches are stitched together with G^n continuity.

The approaches we have presented for geometric continuity are important for several reasons:

- Geometric continuity was previously defined only for first and second order, and then only for curves and surfaces. This work has extended geometric continuity to arbitrary order for splines of arbitrary parametric dimension.
- Geometric continuity for curves and surfaces has been unified in the sense

that they are s two manifestations of the same underlying theory. Previous appro e treated curves and surfaces separately.

- The chain rule n of the Beta constraints is much simpler than the geometry-base ons that have previously been used.

Several applica eometric continuity and the Beta constraints were also presented. These in e placement of Bézier control vertices, the derivation of the cubic Beta-s ve technique, and the construction of a new surface technique: the trian bic Beta-spline.

The triangular eta-spline is a geometrically continuous analog of the triangular cubic B-s s such, it is a G^1 technique, and therefore guarantees tangent plane conti tween spline patches. It possess one shape parameter, has local control, the convex hull property for positive values of the shape parameter. uation algorithm based on recursive subdivision was also developed.

The foundatiometric continuity has now been laid, but this is only the first step toward th flexible, effective, geometrically continuous techniques in a practical CAG m. A comprehensive study of the uses of geometric continuity should n dertaken. Such a study should address the following issues:

- *Shape Functionication.* The Beta constraints provide for the introduction of a pletho ape parameters and shape functions. So many in fact, that it is unrea to expect a viable CAGD system to implement all of them. For insta is possible for a G^3 surface technique to have 6 shape functions per pa undary. This is far too many for a designer to manage effectively. Thu important to determine which of the possible shape functions, or com ons thereof, are useful in a CAGD context.
- *Shape Function cation.* Once non-intuitive shape functions are eliminated, the proble hape function specification must be addressed. Methods must be devel o make it easy for a designer to specify the shape function that is to be yed along a given boundary. Moreover, the method should automatical uarantee compatibility of shape sets, as discussed in Section 6.9.1.
- *Beta-spline Curves.* pls of arbitrary order have been used for sometime in CAGD, and the cu h-spline has recently shown exceptional promise. It is therefore reasona pect that the Beta-splines of arbitrary order are also useful. Although i s to define Beta-splines of arbitrary order in an

abstract mathematical sense, currently a general evaluation algorithm does not exist. Algorithms for subdivision and knot insertion are also desirable, but unknown.

- *Triangular Beta-spline Surfaces.* The triangular cubic Beta-spline was introduced in Chapter 5. However, the situation for triangular Beta-splines is much the same as for Beta-spline curves: construction is still an *ad hoc* process. Thus, algorithms for evaluation, subdivision, and knot insertion for triangular Beta-splines of arbitrary order are essential for the continued study and eventual use of this technique.
- *Multivariate Beta-splines.* The triangular Beta-splines are defined only over a regular triangulation of the domain manifold (the parameter space). In many "real world" design applications, this restriction is too confining. What is needed is a geometrically continuous surface technique defined on a arbitrary tesselation of a domain manifold. Recent work in multivariate B-splines is sufficiently general to describe parametrically continuous surfaces on arbitrary tesselations of a parameter plane [22,45]. The extension to geometric continuity is likely to be non-trivial.

Finally, it may be interesting and fruitful to explore further the notion of a spline as an immersion of a manifold into Euclidean space. The link to manifold theory allows smooth splines of arbitrary topology to be constructed in an inherently coordinate free way, unifies the development of curves, surfaces, volumes, etc., and establishes a framework upon which future applications may be built. These advantages have not come without a price, however. A good deal of mathematical sophistication is required to navigate through the new formalism, but once one acquires "sea legs", the trip can be quite "smooth".

References

1. Alfred V. Aho, John E. Hopcroft, and Jeffery D. Ullman, *The Design and Analysis of Computer Algorithms,* Addison-Wesley, Reading Massachusetts (1979).

2. Robert E. Barnhill, "Representation and Approximation of Surfaces," pp. 68-119 in *Mathematical Software III,* ed. J.R. Rice, Academic Press, New York (1977).

3. Brian A. Barsky, *The Beta-spline: A Local Representation Based on Shape Parameters and Fundamental Geometric Measures*, Ph.D. Thesis, University of Utah, Salt Lake City, Utah (December, 1981).

4. Brian A. Barsky, *Computer Graphics and Computer Aided Geometric Design Using Beta-splines*, Springer-Verlag, Tokyo. To appear.

5. Brian A. Barsky and John C. Beatty, "Local Control of Bias and Tension in Beta-splines," *ACM Transactions on Graphics,* Vol. 2, No. 2, April 1983, pp. 109-134. Also published in *SIGGRAPH '83 Conference Proceedings* (Vol. 17, No. 3), ACM, Detroit, 25-29 July, 1983, pp. 193-218.

6. Brian A. Barsky and Tony D. DeRose, *Geometric Continuity for Parametric Curves,* Technical Report No. UCB/CSD 84/205, Computer Science Division, Electrical Engineering and Computer Sciences Department, University of California, Berkeley, California, USA (October, 1984).

7. Brian A. Barsky, Tony D. DeRose, and Mark D. Dippé, "An Adaptive Subdivision Method with Crack Prevention for Rendering Beta-spline Objects," Submitted for publication.

8. Richard H. Bartels, John C. Beatty, and Brian A. Barsky, *An Introduction to the Use of Splines in Computer Graphics,* UCB/CSD 83/136, Computer Science Division, Electrical Engineering and Computer Sciences Department, University of California, Berkeley, California, USA (August, 1983). Also Tech. Report No. CS-83-9, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada.

9. Richard H. Bartels and John C. Beatty, *Beta-splines with a Difference,* Technical Report No. CS-84-40, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada (May, 1984).

10. Wolfgang Böhm, "Subdividing Multivariate Splines," *Computer-Aided Design,* Vol. 15, No. 6, November, 1983, pp. 345-352.

11. Carl deBoor, "On Calculating with B-splines," *Journal of Approximation Theory,* Vol. 6, No. 1, 1972, pp. 50-60.

12. Carl deBoor, "Splines as Linear Combinations of B-splines. A Survey," pp. 1-47 in *Approximation Theory II,* ed. G. G. Lorentz, C. K. Chui, and L. L. Schumaker, Academic Press, New York (1976).

13. Carl deBoor *A Practical Guide to Splines,* Applied Mathematical Sciences, Vol. 27, Springer-Verlag (1978).

14. William M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry,* Academic Press, New York (1975).

15. F. Brickell and R. S. Clark, *Differentiable Manifolds: An Introduction,* Van Nostrand Rienhold Company, New York (1970).

16. R. C. Buck, *Advanced Calculus,* McGraw-Hill Book Company, Inc., New York (1956).

17. P. de Casteljau, *Courbes et Surfaces à Pôles,* A. A. André Citroen, Paris.

18. Edwin E. Catmull and Raphael J. Rom, "A Class of Local Interpolating Splines," pp. 317-326 in *Computer Aided Geometric Design,* ed. Robert E. Barnhill and Richard F. Riesenfeld, Academic Press, New York (1974).

19. Elaine Cohen, Tom Lyche, and Richard F. Riesenfeld, "Discrete B-splines and Subdivision Techniques in Computer-Aided Geometric Design and Computer Graphics," *Computer Graphics and Image Processing,* Vol. 14, No. 2, October, 1980, pp. 87-111. Also Technical Report No. UUCS-79-117, Department of Computer Science, University of Utah, October 1979.

20. Mauric G. Cox, *The Numerical Evaluation of B-splines,* Report No. NPL-DNACS-4, Division of Numerical Analysis and Computing, National Physical Laboratory, Teddington, Middlesex, England (August, 1971). Also in *J. Inst. Maths. Applics.,* Vol. 10, 1972, pp. 134-149.

21. H. B. Curry and I. J. Schoenberg, "On Spline Distributions and their Limits: The Polya Distribution Functions, Abstract 380t," *Bulletin of the American Mathematical Society,* Vol. 53, p. 1114, 1947.

22. Wolfgang Dahmen and Charles Micchelli, "Multvariate Splines - A New Constructive Approach," pp. 191-215 in *Surfaces in CAGD,* ed. Robert E. Barnhill and Wolfgang Böhm, North-Holland Publishing Company, Amsterdam (1983).

23. Tony D. DeRose and Brian A. Barsky, "Geometric Continuity and Shape Parameters for Catmull-Rom Splines (Extended Abstract)," pp. 57-62 in *Proceedings of Graphics Interface '84,* Ottawa (27 May - 1 June, 1984).

24. Tony D. DeRose and Brian A. Barsky, "An Intuitive Approach to Geometric Continuity for Parametric Curves and Surfaces," pp. 343-351 in *Proceedings of Graphics Interface '85,* Montreal (27-31 May, 1985). Extended abstract in *Proceedings of the International Conference on Computational Geometry and Computer-Aided Design,* New Orleans (5-8 June, 1985), pp. 71-75.

25. Mark E. Dippé, *Antialiasing in Computer Graphics,* Ph.D. Thesis, University of California, Berkeley, California (April, 1985).

26. Manfred P. DoCarmo, *Differential Geometry of Curves and Surfaces,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1976).

27. Gerald Farin, "Visually C^2 Cubic Splines," *Computer-Aided Design,* Vol. 14, No. 3, May, 1982, pp. 137-139.

28. Gerald Farin, "Designing C^1 Surfaces Consisting of Triangular Cubic Patches," *Computer-Aided Design,* Vol. 14, No. 5, September, 1982, pp. 253-256.

29. Gerald Farin, "A Construction for Visual C^1 Continuity of Polynomial Surface Patches," *Computer Graphics and Image Processing,* Vol. 20, 1982, pp. 272-282.

30. Gerald Farin, "Smooth Interpolation to Scattered 3D Data," pp. 43-63 in *Surfaces in CAGD,* ed. Robert E. Barhhill and Wolfgang Böhm, North-Holland Publishing Company, Amsterdam (1983).

31. Richard J. Fateman, *Addendum to the MACSYMA Reference Manual for the VAX,* Computer Science Division, University of California, Berkeley (1982).

32. Ivor D. Faux and Michael J. Pratt, *Computational Geometry for Design and Manufacture,* Ellis Horwood Ltd. (1979).

33. Alain Fournier and Brian A. Barsky, "Geometric Continuity with Interpolating Bézier Curves (Extended Summary)," pp. 337-341 in *Proceedings of Graphics Interface '85,* Montreal (27-31 May, 1985).

34. A. H. Fowler and C. W. Wilson, "Cubic Spline, A Curve Fitting Routine," Union Carbide Corporation Report, Y-1400 (Rev. I.) (1966).

35. Steven A. Gabriel and James T. Kajiya, "Spline Interpolation in Curved Manifolds," Submitted for publication.

36. Ronald N. Goldman, "Subdivision Algorithms for Bézier Triangles," *Computer-Aided Design,* Vol. 15, No. 3, May, 1983, pp. 159-166.

37. T. N. T. Goodman, "Properties of β-Splines," accepted for publication in *Journal of Approximation Theory,* Vol. 43, 1985.

38. William J. Gordon, "Spline-Blended Surface Interpolation through Curve Networks," *Journal of Mathematics and Mechanics,* Vol. 18, No. 19, 1969, pp. 931-952. Also Research Publication GMR-921, General Motors Research Laboratories, September 1969.

39. William J. Gordon and Richard F. Riesenfeld, "B-spline Curves and Surfaces," pp. 95-126 in *Computer Aided Geometric Design,* ed. Robert E. Barnhill and Richard F. Riesenfeld, Academic Press, New York (1974).

40. Martin L. Griss, *A REDUCE Symbolic-Numeric Tutorial,* Utah Symbolic Computation Group Operating Note, Technical Report No. UCP-32, Department of Computer Science, University of Utah (October, 1977).

41. Victor Guillemin and Alan Pollack, *Differential Topology,* Prentice-Hall, Inc., New Jersey (1974).

42. Anthony C. Hearn, "REDUCE: A User-Oriented Interactive System for Algebraic Simplification," in *Interactive Systems for Experimental Applied Mathematics,* ed. M. Klerer and J. Reinfelds Academic Press, New York (1968).

43. Klaus Höllig, "Multivariate Splines," *SIAM Journal of Numerical Analysis,* Vol. 19, No. 5, October, 1982, pp. 1013-1031.

44. Juergen Kahmann, "Continuity of Curvature Between Adjacent Bézier Patches," pp. 65-75 in *Surfaces in Computer Aided Geometric Design,* ed.

Robert E. Barnhill and Wolfgang Boehm, North-Holland Publishing Company (1983).

45. Peter Kochevar, "An Application of Multivariate B-splines to Computer-Aided Geometric Design," *Rocky Mountain Journal of Mathematics,* Vol. 14, No. 1, Winter, 1974, pp. 159-174.

46. Jeffrey M. Lane, Loren C. Carpenter, J. Turner Whitted, and James F. Blinn, "Scan Line Methods for Displaying Parametrically Defined Surfaces," *Communications of the ACM,* Vol. 23, No. 1, January, 1980, pp. 23-34.

47. J. R. Manning, "Continuity Conditions for Spline Curves," *The Computer Journal,* Vol. 17, No. 2, May, 1974, pp. 181-186.

48. Frederick Munchmeyer, "Mathematical Ship Lines and Surfaces," *Marine Technology,* Vol. 19, No. 3, July, 1982, pp 219-227.

49. Gregory M. Nielson, "Some Piecewise Polynomial Alternatives to Splines under Tension," pp. 209-235, in *Computer Aided Geometric Design,* ed. Robert E. Barnhill and Richard F. Riesenfeld, Academic Press, New York (1974).

50. Lyle Ramshaw, "A Euclidean View of Joints between Bézier Curves." Submitted for publication.

51. Richard F. Riesenfeld, *Applications of B-spline Approximation to Geometric Problems of Computer-Aided Design,* Ph.D. Thesis, Syracuse University (May, 1973). Available as Tech. Report No. UTEC-CSc-73-126, Department of Computer Science, University of Utah.

52. Malcolm A. Sabin, *Parametric Splines in Tension,* Technical Report No. VTO/MS/160, British Aircraft Corporation, Weybridge, Surrey, England (July 23, 1970).

53. Malcolm A. Sabin, *The Use of Piecewise Forms for the Numerical Representation of Shape,* Ph.D. Thesis, Budapest (1976).

54. Daniel G. Schweikert, "An Interpolation Curve Using a Spline in Tension," *Journal of Mathematics and Physics,* Vol. 45, 1966, pp. 312-317.

55. Thomas W. Sederberg, *Implicit and Parametric Curves and Surfaces for Computer Aided Geometric Design,* Ph.D. Thesis, Purdue University (1983).

56. Ken Shoemake, "Animating Rotation with Quaternion Curves," pp. 245-254 in *SIGGRAPH '85 Conference Proceedings,* ACM, San Francisco, (July, 1985).

57. Michael Spivak, *Calculus on Manifolds,* W. A. Benjamin, Inc., New York (1965).

58. Michael Spivak, *A Comprehensive Introduction to Differential Geometry, Vol. I,* Publish or Perish, Boston (1970).

59. J. L. Synge and A. Schild, *Tensor Calculus,*. Dover Publications, New York (1978).

60. M. Veron, G. Ris, and J.-P. Musse, "Continuity of Biparametric Surface Patches," *Computer-Aided Design,* Vol. 8, No. 4, October, 1976, pp. 267-273.